LA-UR--97-31

HUMAN RETROVIRUSES and AIDS 1996

A COMPILATION AND ANALYSIS OF

NUCLEIC ACID AND AMINO ACID SEQUENCES

Editors

Gerald Myers
Theoretical Division
Los Alamos National Laboratory

Bette Korber
Theoretical Division
Los Alamos National Laboratory

Brian Foley
Theoretical Division
Los Alamos National Laboratory

Kuan-Teh Jeang
Molecular Virology Section
National Institutes of Health

John W. Mellors
Graduate School of Public Health/VAMC
University of Pittsburgh

Simon Wain-Hobson
UREG, Pasteur Institute

Database and Analysis Staff

Sampath Billikanti, John Blouin, Charles Calef,
Esther Guzman, Hong Lu, Kersti MacInnes

Los Alamos National Laboratory

This publication is being funded by the Division of AIDS, National Institute of Allergy and Infectious Diseases, through an interagency agreement with the U.S. Department of Energy.

Published by Theoretical Biology and Biophysics
Group T-10, Mail Stop K710
Los Alamos National Laboratory, Los Alamos, New Mexico 87545 U.S.A.

LA-UR 97-31

DISCLAIMER

CONTENTS

More detailed tables of contents are to be found within the various parts of the compendium. The following provides merely an overview.

ACKNOWLEDGMENTS

The HIV Sequence Database and Analysis Project is funded by the Vaccine and Prevention Research Program of the AIDS Division of the National Institute of Allergy and Infectious Diseases (Dr. James Bradac, Project Officer) through an interagency agreement with the U.S. Department of Energy.

We thank the many researchers who have made their sequences available prior to publication.

We also thank last year's editors, Beatrice Hahn, Francine McCutchan, Lou Henderson and George Pavlakis, who were not a part of this year's publication but who may, at one time or another, return in this role.

INTRODUCTION

This compendium and the accompanying floppy diskettes are the result of an effort to compile and rapidly publish all relevant molecular data concerning the human immunodeficiency viruses (HIV) and related retroviruses. The scope of the compendium and database is best summarized by the five parts that it comprises: (I) Nucleic Acid Alignments and Sequences; (II) Amino Acid Alignments; (III) Analysis; (IV) Related Sequences; and (V) Database Communications. Information within all the parts is updated throughout the year on the Web site, *http://hiv-web.lanl.gov*.

While this publication could take the form of a review or sequence monograph, it is not so conceived. Instead, the literature from which the database is derived has simply been summarized and some elementary computational analyses have been performed upon the data. Interpretation and commentary have been avoided insofar as possible so that the reader can form his or her own judgments concerning the complex information. In addition to the general descriptions below of the parts of the compendium, the user should read the individual introductions for each part.

Part I. Nucleic Acid Alignments and Sequences. Annotated nucleic acid sequences of certain HIVs and SIVs are presented in a form close to that of the GenBank Sequence Library. Our few modifications of standard GenBank format were instituted to better serve the particular community for which this database is intended. Beginning in 1995, most sequences are not presented but rather are catalogued, in order to conserve space; the full formatted GenBank entries of these sequences are located on the Web site (*http://hiv-web.lanl.gov*) and the database FTP server.

The LOCUS name or identifier of an entry may differ slightly from that found in the GenBank or EMBL libraries, but the ACCESSION numbers are identical for entries in all (four) nucelotide sequence databases. Thus each entry is universally and uniquely traceable. Sequences may also be described by COMMON NAMES.The SOURCE line provides information, when available, about the infectivity or biological activity of the molecular clone from which a sequence has been derived. REFERENCES are limited to literature or personal communications having authority for the original sequence data; references that review sequence information, or that shed light upon the function or variation of coding and regulatory sequences, are listed in Part V.

Entries in Part I are annotated within the sequence, while their GenBank or EMBL-formatted versions on the floppy diskettes make use of FEATURES tables. The hard-copy annotation includes coding regions, regulatory structures, splice sites, and other features of functional significance. The authority for this annotation is largely invariance, the recurrence of patterns such as TATAA and AATAAA. Although our practice has been to conservatively annotate, we caution the user against docility: sequence information regarding transcripts, for example, is far from certain or complete at this time. Part I is sometimes divided into three subsections, A, B, and C, concerned with HIV-1s, HIV-2/SIVs, and SIVAGMs, SIVMND and SIVSYK.

Part II. Amino Acid Alignments. This section contains in alignment the amino acid sequences (mostly full-length) of all known coding regions and open reading frames of HIV-1, HIV-2, and SIVs. In 1996, large alignments are presented in terms of representative primate immunodeficiency viral types and subtypes. Consensus sequences for HIV-1 subtypes have also been aligned in this release. Other alignments are available on the Web site. Protein processing sites are annotated when known. The reader should consult the introduction to Part II for further explanation of the presentation and annotation of the amino acid sequences.

Part III. Analysis. This section is open-ended with the constraint that the sequence analyses and compilations be basic and of interest to the diversity of users. In 1996, analyses and curatorial contributions include tables of mutations relating to drug resistance, sections on Tat and VpU, analyses of sequence alignment and structure prediction, discussions of HIV-1 subtypes, etc.

Part IV. Related Sequences. Heretofore, this section of the compendium has featured HIV related viral sequences—of nonprimate lentiviruses and the human T-cell lymphotropic viruses. Beginning in 1993, Part IV entries include, with greater emphasis, coding sequences for cellular proteins involved with HIV pathogenesis. In 1996, there are articles summarizing cellular factors and viral mimicry.

Part V. Database Communications. This part contains a printed supplemental reference list for citations in 1995 and 1996. It also provides diskettes of sequences and information about accessing sequences through Internet. The floppy diskettes contain new nucleic acid sequences from Part I (especially those found in alignments) and Part IV and their translated amino acid sequences. For the most current information regarding database files, see the READ.ME file on each diskette. Nucleotide entries are presented in GenBank format for North American users and in EMBL format for European users (unless otherwise requested). Similarly, amino acid sequences are in either PIR or Swiss-Prot format. The diskettes themselves are either 3.5" IBM-DOS format or 3.5" Macintosh format, depending upon what has been requested. If there is any trouble using these files with software designed to work with the format we have sent, please let us know the name of the program you are using and the file that it could not handle.

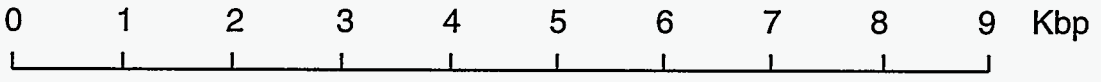
A comprehensive compilation of the nucleic acid and protein sequences published in the Human Retroviruses and AIDS Database since 1987 is available through our Web site, *http://hiv-web.lanl.gov* and on our FTP Server, as described in Part V.

We are prepared to quickly enter both protein and nucleotide sequences into the Human Retroviruses and AIDS database, and in the case of nucleotide sequences, oversee their entry into the large gene libraries. Submission of unpublished sequences is invited and encouraged. Sequence data or inquiries regarding the database should be addressed to

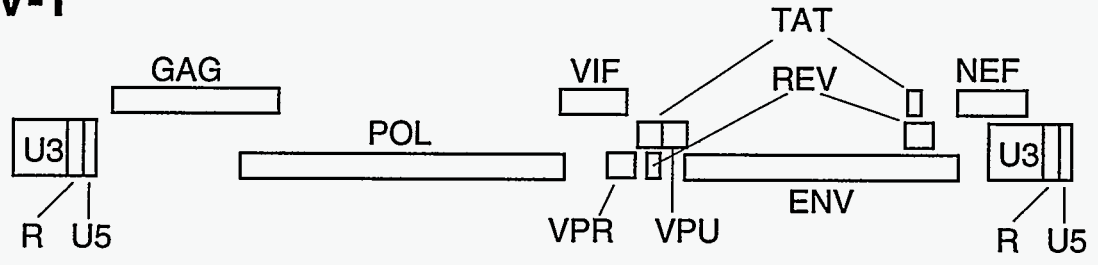
Gerald Myers
Theoretical Division
T-10, MS K710
LANL
Los Alamos, NM 87545

(505)-665-0480; fax (505)-665-3493
e-mail: glm@t10.lanl.gov

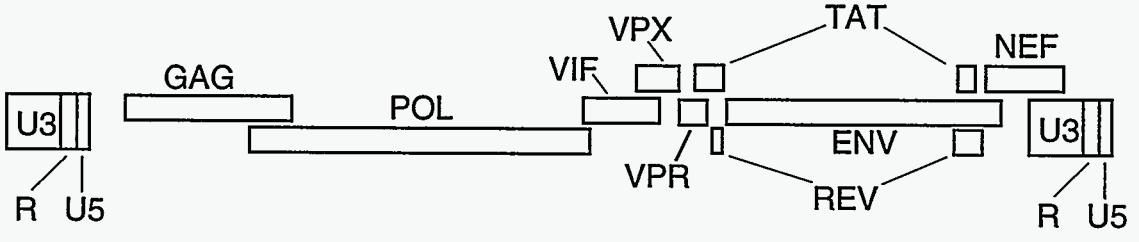
A short glossary follows.


0
1
2
3
4
5
6
7
8
9
Kbp

HIV-1


TAT
GAG
VIF
REV
NEF
U3
POL
U3
ENV
R
U5
VPR
VPU
R
U5

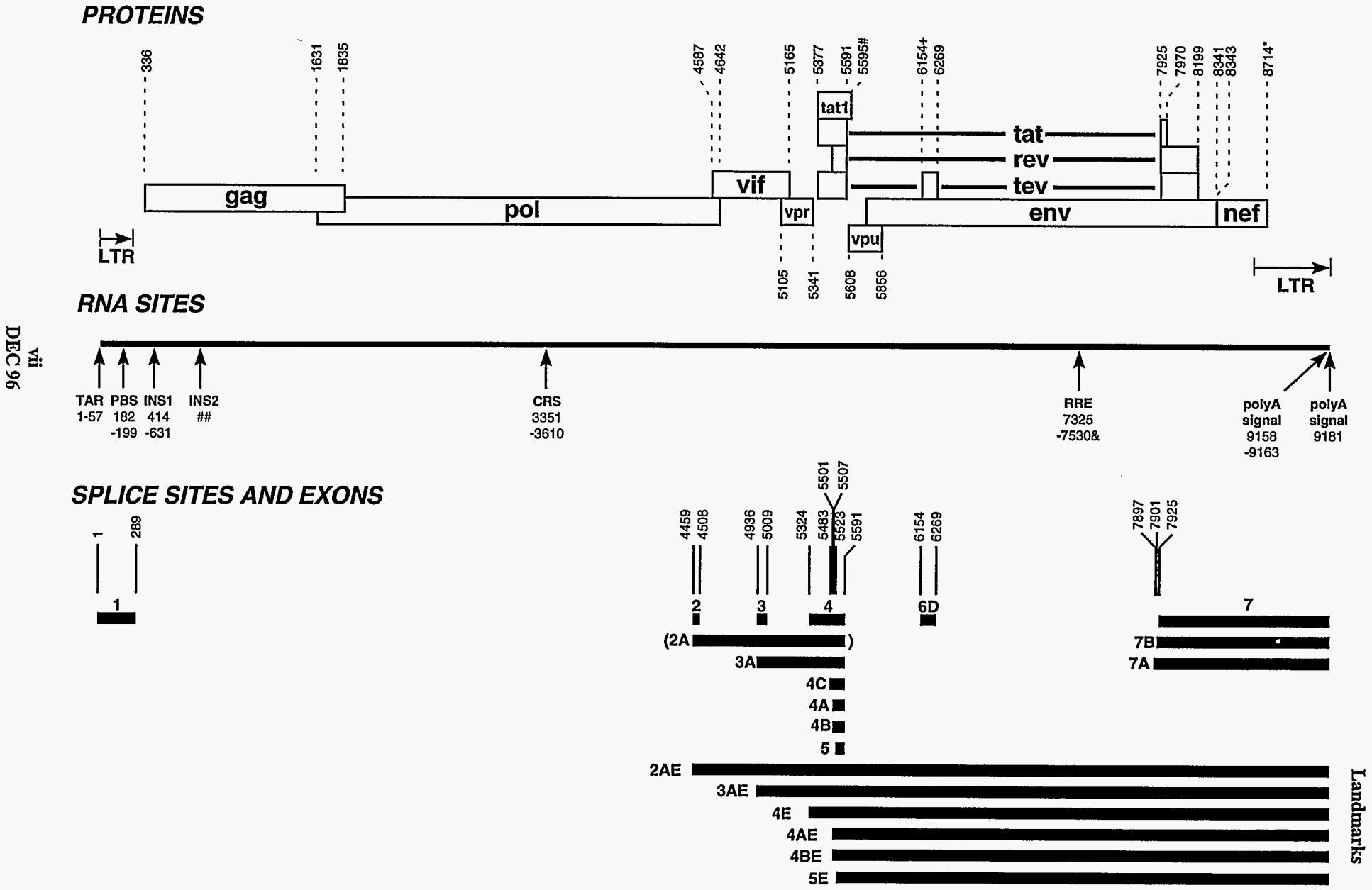
HIV-2


VPX
TAT
NEF
VIF
GAG
U3
POL
ENV
U3
VPR
REV
R
U5
R
U5

HIV/SIV PROTEINS

NAME	SIZE	FUNCTION	LOCALIZATION
Gag MA	p17	membrane anchoring; env interaction; nuclear transport of viral core. (myristylated protein)	virion
CA	p24	core capsid	virion
NC	p7	nucleocapsid, binds RNA	virion
	p6	binds Vpr	virion
Protease (PR)	p15	gag/pol cleavage and maturation	virion
Reverse transcriptase (RT), RNase H	p66 p51 (heterodimer)	reverse transcription, RNase H activity	virion
Integrase (IN)		DNA provirus integration	virion
Env	gp120/gp41	external viral glycoproteins bind to CD4 receptor	plasma membrane, virion envelope
Tat	p16/p14	viral trancriptional transactivator	primarily in nucleolus/nucleus
Rev	p19	RNA transport, stability and utilization factor (phosphoprotein)	primarily in nucleolus/nucleus shuttling between nucleolus and cytoplasm
Vif	p23	promotes virion maturation and infectivity	cytoplasm (cytosol, membranes) virion
Vpr	p10-15	promotes nuclear localization of preintegration complex, inhibits cell division, arrests infected cells at G2/M	virion, nucleus (nuclear membrane?)
Vpu	p16	promotes extracellular release of viral particles; degrades CD4 in the ER; (phosphoprotein only in HIV-1 and SIVcpz)	integral membrane protein
Nef	p27-p25	CD4 downregulation (myristylated protein)	plasma membrane, cytoplasm (virion?)
Vpx	p12-16	vpr homolog? (not in HIV-1, only in HIV-2 and SIV)	virion (nucleus?)
Tev	p28	tripartite tat-env-rev protein (also named Tnv)	primarily in nucleolus/nucleus

LANDMARKS ON THE HIV-1 GENOMIC RNA



DEC 96 vii

Landmarks

LANDMARKS:

HIV GENOMIC STRUCTURAL ELEMENTS

LTR Long terminal repeat, the DNA sequence flanking the genome of integrated proviruses. It contains important regulatory regions, especially those for transcription initiation and polyadenylation.

TAR Target sequence for viral transactivation, the binding site for Tat protein and for cellular proteins; consists of approximately the first 45 nucleotides of the viral mRNAs in HIV-1 (or the first 100 nucleotides in HIV-2 and SIV.) TAR RNA forms a hairpin stem-loop structure with a side bulge; the bulge is necessary for Tat binding and function.

RRE Rev responsive element, an RNA element encoded within the env region of HIV-1. It consists of approximately 200 nucleotides (positions 7327 to 7530 from the start of transcription in HIV-1.) The RRE is necessary for Rev function; it contains a high affinity site for Rev; in all, approximately seven binding sites for Rev exist within the RRE RNA. Other lentiviruses (HIV-2, SIV, visna, CAEV) have similar RRE elements in similar locations within env, while HTLVs have an analogous RNA element (RXRE) serving the same purpose within their LTR; RRE is the binding site for Rev protein, while RXRE is the binding site for Rex protein. RRE (and RXRE) form complex secondary structures, necessary for specific protein binding.

CRS cis-acting repressive sequences postulated to inhibit structural protein expression in the absence of Rev. One such site was mapped within the pol region of HIV-1. The exact function has not been defined; splice sites have been postulated to act as CRS sequences.

INS Inhibitory/Instability RNA sequences found within the structural genes of HIV-1 and of other complex retroviruses. Multiple INS elements exist within the genome and can act independently; one of the best characterized elements spans nucleotides 414 to 631 in the gag region of HIV-1. The INS elements have been defined by functional assays as elements that inhibit expression posttranscriptionally. Mutation of the RNA elements was shown to lead to INS inactivation and up regulation of gene expression.

GENES AND GENE PRODUCTS

GAG genomic region encoding the capsid proteins (group specific antigens). The precursor is the p55 myristylated protein, which is processed to p17 (MAtrix), p24 (CApsid), p7 (NucleoCapsid), and p6 proteins, by the viral protease. Gag associates with the plasma membrane where the virus assembly takes place. The 55 kDa Gag precursor is called assemblin to indicate its role in viral assembly.

POL the genomic region encoding the viral enzymes protease, reverse transcriptase and integrase. These enzymes are produced as a Gag-pol precursor polyprotein, which is processed by the viral protease; the Gag-pol precursor is produced by ribosome frameshifting at the C-terminus of gag.

ENV viral glycoproteins produced as a precursor (gp160) and processed to the external glycoprotein gp120 and the transmembrane glycoprotein gp41. The mature proteins are held together by non-covalent interactions; as a result, a substantial amount of gp120 is released in the medium. gp120 contains the binding site for the CD4 receptor.

TAT Transactivator of HIV gene expression. One of the two necessary viral regulatory factors (Tat and Rev) for HIV gene expression. Two forms are known, Tat-1exon (minor form) of 72 amino acids and Tat-2exon (major form) of 86 amino acids. The electrophoretic mobility of these two forms in SDS gels is anomalous, with apparent sizes of approximately 16 kD and 14 kD for Tat- 2exon and Tat-1exon, respectively. Low levels of both proteins are found in persistently infected cells. Tat has been localized primarily in the nucleolus/nucleus by immunofluorescence. It acts by binding to the TAR RNA element and activating transcription initiation and/or elongation from the LTR promoter. It is the first eukaryotic transcription factor known to interact with RNA rather

than DNA and may have similarities with prokaryotic anti- termination factors. Extracellular Tat can be found and can be taken up by cells in culture.

REV The second necessary regulatory factor for HIV expression. A 19 kD phosphoprotein, localized primarily in the nucleolus/nucleus, Rev acts by binding to RRE and promoting the nuclear export, stabilization and utilization of the viral mRNAs containing RRE. Rev is considered the most functionally conserved regulatory protein of lentiviruses. Rev cycles rapidly between the nucleus and the cytoplasm.

VIF Viral infectivity factor, typically 23 kD. Promotes the infectivity but not the production of viral particles. In the absence of Vif the produced viral particles are defective, while the cell-to-cell transmission of virus is not affected significantly. Found in almost all lentiviruses, Vif is a cytoplasmic protein, existing in both a soluble cytosolic form and a membrane-associated form. The latter form of Vif is a peripheral membrane protein that is tightly associated with the cytoplasmic side of cellular membranes. Some recent observations suggest that Vif is incorporated in the virion.

VPR Vpr (viral protein R) is a 96-amino acid (14 kd) protein, which is incorporated into the virion. It interacts with the p6gag part of the Pr55gag precursor. Vpr detected in the cell is localized to the nucleus. Proposed functions for Vpr include the nuclear import of preintegration complexes, cell growth arrest, transactivation of cellular genes, and induction of cellular differentiation. Found in HIV-1, HIV-2, SIVmac and SIVmnd. It is homologous to VPX of SIVagm.

VPU Vpu (viral protein U) is unique to HIV-1 and SIVcpz, a close relative of HIV-1. There is no similar gene in HIV-2 or SIV. Vpu is a 16-kd (81-amino acid) type I integral membrane protein with at least two different biological functions: (a) degradation of CD4 in the endoplasmic reticulum, and (b) enhancement of virion release from the plasma membrane of HIV-1-infected cells. Vpu probably possesses an N-terminal hydrophobic membrane anchor and a hydrophilic moiety. It is phosphorylated by casein kinase II at positions Ser52 and Ser56. Vpu is involved in env maturation; not found in the virion.

NEF (previously named 3′ ORF) is an approximately 27-kd myristylated protein produced by an ORF located at the 3′ end of the primate lentiviruses. Other forms of Nef are known, including nonmyristylated variants. Nef is predominantly cytoplasmic and associated with the plasma membrane via the myristyl residue linked to the conserved second amino acid (Gly). Nef has also been identified in the nucleus and found associated with the cytoskeleton in some experiments. Its association with the virion is suspected but not proven. One of the first HIV proteins to be produced in infected cells, it is the most immunogenic of the accessory proteins. Initially thought to be a negative factor, Nef was found to be important for viral replication in vivo. The nef genes of HIV and SIV are dispensable in vitro, but are essential for efficient viral spread and disease progression in vivo. Nef is necessary for the maintenance of high virus loads and for the development of AIDS in macaques. Nef downregulates CD4, the primary viral receptor, and is also proposed to increase viral infectivity. Nef contains PxxP motifs that bind to SH3 domains of a subset of Src kinases and are required for the enhanced growth of HIV but not for the downregulation of CD4.

VPX Virion protein of 12 kD found only in HIV-2/SIVagm and not in HIV-1 or SIVmnd. Vpx function in relation to Vpr is not fully elucidated. Vpx is necessary for efficient replication of SIV in PBMCs. Some studies indicate that Vpx and Vpr proteins may be functionally distinct. Progression to AIDS and death in SIV-infected animals can occur in the absence of Vpr or Vpx. Double mutant virus lacking both vpr and vpx was severely attenuated, whereas the single mutants were not, suggesting a redundancy in the function of Vpr and Vpx related to virus pathogenicity.

TEV (also named tnv) tripartite 28 kD viral phosphoprotein produced by some HIV-1 strains. Found primarily in the nucleolus/nucleus. Tev contains the first exon of Tat, a small part of Env and the second exon of Rev. It exhibits both Tat and Rev functions and can functionally replace both essential regulatory proteins of HIV-1. It is produced very early in infection.

STRUCTURAL PROTEINS/VIRAL ENZYMES The products of gag, pol and env genes, which are essential components of the retroviral particle.

REGULATORY PROTEINS Tat and Rev proteins of HIV/SIV and Tax and Rex proteins of HTLVs. They modulate transcriptional and posttranscriptional steps of virus gene expression and are essential for virus propagation.

ACCESSORY OR AUXILIARY PROTEINS additional virion and non-virion- associated proteins produced by HIV/SIV retroviruses: Vif, Vpr, Vpu, Vpx, Nef. Although the accessory proteins are in general not necessary for viral propagation in tissue culture, they have been conserved in the different isolates; this conservation and experimental observations suggest that their role in vivo is very important.

COMPLEX RETROVIRUSES Retroviruses regulating their expression via viral factors and expressing additional proteins (regulatory and accessory) essential for their life cycle.

I

Nucleic Acid Alignments and Sequences

Nucleotide sequence alignments were generated in the 1996 compendium by two different approaches: On the one hand, a Hidden Markov Method (HMM) was used as described in Part III (Myers and Farmer). For simplicity, only representative PIVs are shown; in some cases, hundreds of sequences contributed to the alignment. A simple consensus over just the representative sequences was deduced using MASE. On the other hand, new HIV-1 subtype sequences were added to previous alignments generated using the PIMA and MASE tools, as in earlier compendiums. Alignments of subtype consensus are presented. Mixed case consensus sequences are used as the reference sequences for each alignment. Upper case letters indicate 100% conservation of nucleotide bases in a given position of the alignment, and lower case letters represent bases conserved in at least 50% of the sequences. The symbol "?" indicates no consensus at a position.

With few exceptions, only full-length coding sequences were included in these alignments. Tables of information pertaining to each sequence in an alignment are provided. The common names given to sequences in alignment and in the accompanying tables were selected on several grounds: for sequences corresponding to samples provided by the NIAID repository, WHO and DAIDS conventions are employed for the names; for other sequences, the name given by the authors of the paper reporting the sequence are usually utilized. We wish to thank many sequencers for providing data prior to publication.

Prior to the 1992 database compendium, HIV-1 sequences had been roughly classified as 'U.S.' and 'African.' In light of many new HIV-1 *gag* and *env* sequences, it became more useful, starting with the 1992 compendium, to categorize HIV-1 sequences into five sequence subtypes, depending upon the coding sequence. HIV-1 subtypes now number ten (A through J). Collectively, these are called group M sequences, as they are significantly distinct from group O sequences. The bases for this classification, discussed at greater length in Part III of this compendium, are:

i) subtypes are approximately equidistant from one another in *env* (a "star phylogeny");

ii) the *env* phylogenetic tree is for the most part congruent with *gag* phylogenetic trees;

iii) two or more samples are required to define a sequence subtype.

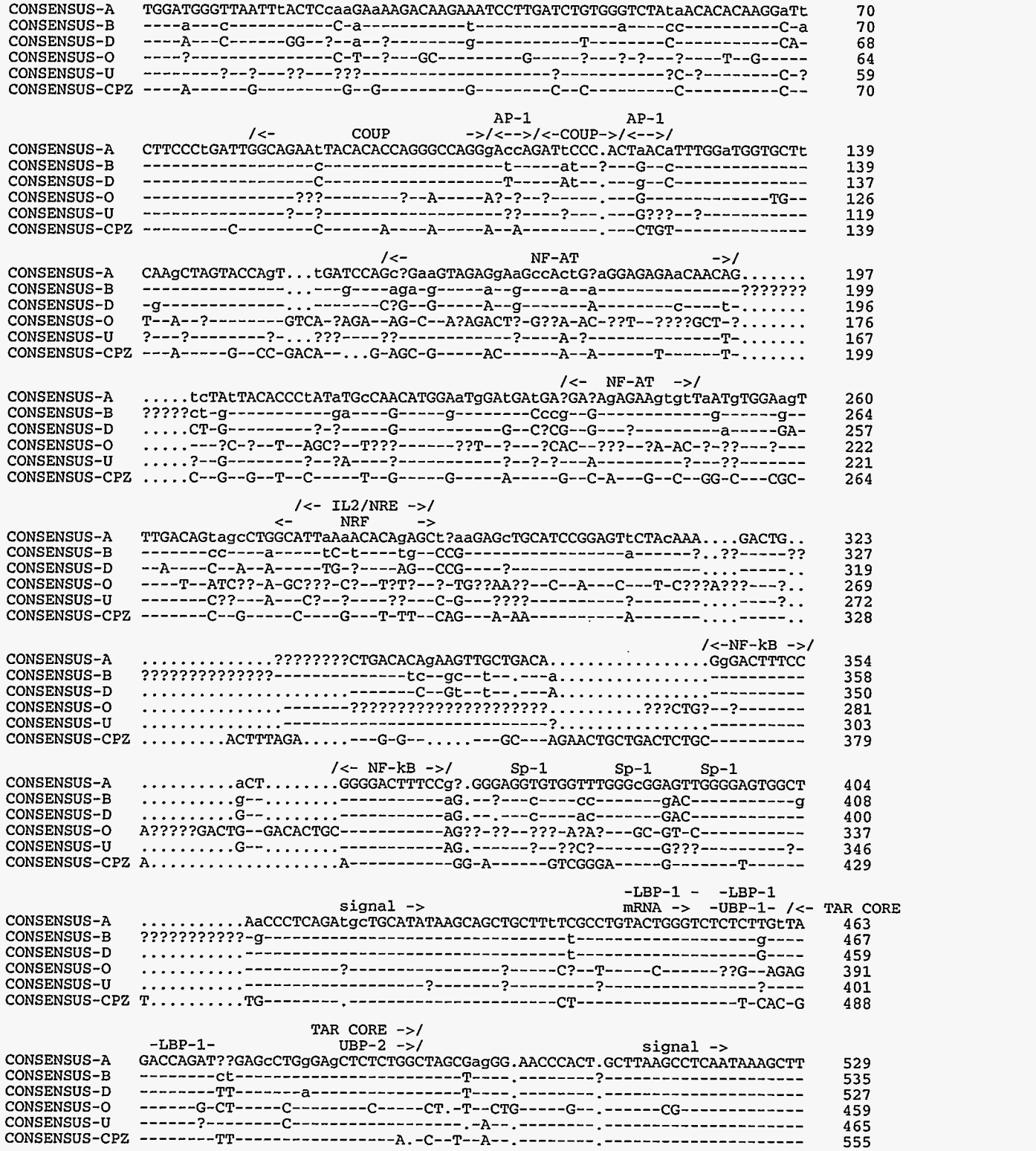
Subtype naming problems have arisen for several reasons. A small but not insignificant number of viral sequences are hybrid, clustering with one sequence subtype in *gag* and another sequence subtype in *env*, for example; or, to take another example, clustering over different stretches with two or more subtypes in *env*. All subtype E sequences based on *env* have *gag* sequences that align with subtype A sequences; and, moreover, the 3′ half of the gp41 cds of E and subtype G *env* sequences align with subtype A sequences (Gao *et al, J. Virol.* **70**:1651,1996). Given the homogeneity of G and E subtype sequences, these are handled as subtypes and not as hybrids in the following alignments. It remains to be discovered whether these states have arisen from recombination or lack of divergence.

Several analyses of mosaic molecules were presented in Part III of the 1995 compendium. Naming also becomes problematic when highly divergent forms of a given subtype arise: such forms are sometimes designated A′, B′, F′ , etc. It is increasingly necessary to have sequence data from both *gag* and *env* coding sequences when a new form or subtype is being claimed.

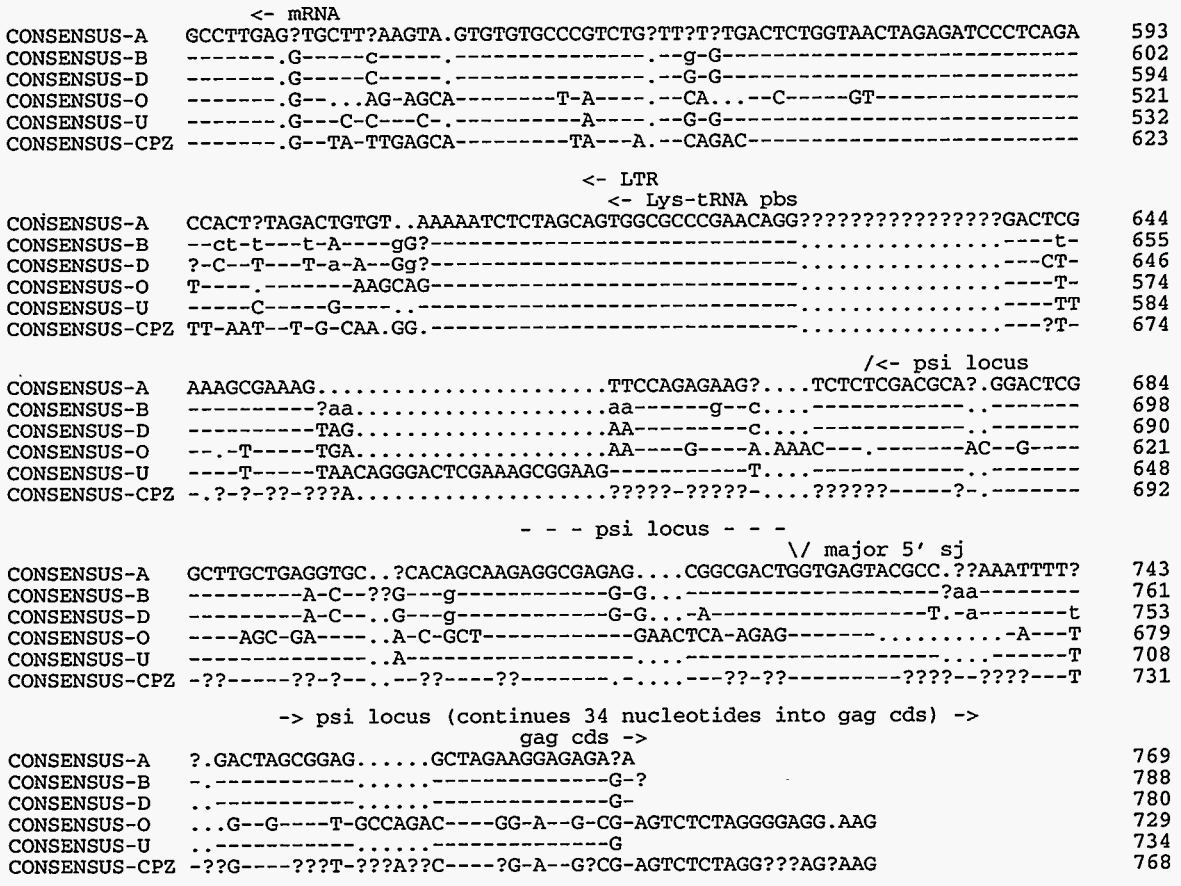
The authority for some of the annotation is limited largely to invariance—the recurrence of patterns such as AATAAA, for example. The reader should be cautious in drawing upon this information. Due to space limitations, only certain sequences are reported as entries in this section. Beginning in 1995, most sequences are cataloged in Part I using just the headers of GenBank entries; to gain the full entry including the sequence itself, users should go to the HIV Sequence Database WWW site (*http://hiv-web.lanl.gov*),

the FTP site (also described in Part V of this compendium) or any of the large gene libraries. Sequences used in alignments are provided on the accompanying diskettes, and the alignments themselves are on the WWW site. Sequence entries from previous years are presented in the 1990–1994 compendiums.

Contents Part I 1996


U3 start ->
CONSENSUS-A TGGATGGGTTAATTtACTCcaaGAaAAGACAAGAAATCCTTGATCTGTGGGTCTAtaACACACAAGGaTt 70
CONSENSUS-B ----a---c----------C-a----------t--------------a----cc----------C-a 70
CONSENSUS-D ----A---C------GG--?--a--?--------g----------T--------C----------CA- 68
CONSENSUS-O ----?---------------C-T--?---GC---------G-----?---?-?---?----T--G----- 64
CONSENSUS-U --------?--?---??---???--------------------?----------?C-?--------C-? 59
CONSENSUS-CPZ ----A------G---------G--G---------G--------C--C---------C----------C-- 70
AP-1 AP-1
/<- COUP ->/<-->/<-COUP->/<-->/
CONSENSUS-A CTTCCCtGATTGGCAGAAtTACACACCAGGGCCAGGgAccAGATtCCC.ACTaACaTTTGGaTGGTGCTt 139
CONSENSUS-B ----------c----------t-----at--?---G--c---------- 139
CONSENSUS-D ----------C----------T-----At--.---g--C---------- 137
CONSENSUS-O ----------???--------?--A-----A?-?--?-----.---G----------TG-- 126
CONSENSUS-U ----------?--?----------??----?---.---G???--?---------- 119
CONSENSUS-CPZ ---------C--------C------A----A-----A--A--------.---CTGT---------- 139
/<- NF-AT ->/
CONSENSUS-A CAAgCTAGTACCAgT...tGATCCAGc?GaaGTAGAGgAaGccActG?aGGAGAGAaCAACAG..... 197
CONSENSUS-B ----------...---g----aga-g-----a--g----at----?-------?????? 199
CONSENSUS-D -g----------...-------C?G--G-----A--g-------A--------c----t-..... 196
CONSENSUS-O T--A--?--------GTCA-?AGA--AG-C--A?AGACT?-G??A-AC-??T--????GCT-?..... 176
CONSENSUS-U ?---?--------?-...???----??----------?----A-?----------T-..... 167
CONSENSUS-CPZ ---A-----G--CC-GACA--...G-AGC-G-----AC------A--A------T------T-..... 199
/<- NF-AT ->/
CONSENSUS-A .....tcTAtTACACCCtATaTGcCAACATGGAaTgGAtGAtGA?GA?AgAGAAgtgtTaATgTGGAagT 260
CONSENSUS-B ?????ct-g----------ga----G-----g--------Cccg--G----------g------g-- 264
CONSENSUS-D .....CT-G---------?-?-----G----------G--C?CG--G---?---------a-----GA- 257
CONSENSUS-O .....---?C-?--T--AGC?--T???------??T--?---?CAC--???--?A-AC-?-??---?--- 222
CONSENSUS-U .....?--G---------?--?A----?----------?--?-?---A---------?---??------- 221
CONSENSUS-CPZ .....C--G--G--T--C-----T--G-----G-----A-----G--C-A---G--C--GG-C---CGC- 264
/<- IL2/NRE ->/
<- NRF ->
CONSENSUS-A TTGACAGtagcCTGGCATTaAaACACAgAGCt?aaGAGcTGCATCCGGAGTtCTAcAAA....GACTG.. 323
CONSENSUS-B -------cc----a-----tC-t----tg--CCG----------a------?..??-----?? 327
CONSENSUS-D --A----C--A--A-----TG-?----AG--CCG----?----------...-----.. 319
CONSENSUS-O ----T--ATC??-A-GC???-C?--T?T?--?-TG??AA??--C--A---C---T-C???A???---?.. 269
CONSENSUS-U -------C??---A---C?--?----??---C-G---????---------?-------...----?.. 272
CONSENSUS-CPZ -------C--G-----C----G---T-TT--CAG---A-AA----------A-------...-----.. 328
/<-NF-kB ->/
CONSENSUS-A ....????????CTGACACAgAAGTTGCTGACA.....GgGACTTTCC 354
CONSENSUS-B ????????????????----------tc--gc--t--.---a.....---------- 358
CONSENSUS-D .....-------C--Gt--t--.---A.....---------- 350
CONSENSUS-O .....-------??????????????????.....???CTG?--?------- 281
CONSENSUS-U .....----------?.....---------- 303
CONSENSUS-CPZ .....ACTTTAGA.....---G-G--.....---GC---AGAACTGCTGACTCTGC---------- 379
/<- NF-kB ->/ Sp-1 Sp-1 Sp-1
CONSENSUS-A .....aCT.....GGGGACTTTCCg?.GGGAGGTGTGGTTTGGGcGGAGTTGGGGAGTGGCT 404
CONSENSUS-B .....g--.....----------aG.--?---c----cc-------gAC----------g 408
CONSENSUS-D .....G--.....----------aG.--.---c----ac-------GAC---------- 400
CONSENSUS-O A?????GACTG--GACACTGC----------AG??-??--???-A?A?---GC-GT-C---------- 337
CONSENSUS-U .....G--.....----------AG.------?--??C?-------G???---------?- 346
CONSENSUS-CPZ A.....A----------GG-A------GTCGGGA-----G-------T------ 429
-LBP-1 - -LBP-1
signal -> mRNA -> -UBP-1- /<- TAR CORE
CONSENSUS-A .....AaCCCTCAGAtgcTGCATATAAGCAGCTGCTTtTCGCCTGTACTGGGTCTCTCTTGtTA 463
CONSENSUS-B ???????????-g----------t----------g---- 467
CONSENSUS-D .....----------t----------G---- 459
CONSENSUS-O .....----------?----------?-----C?--T-----C------??G--AGAG 391
CONSENSUS-U .....----------?-------?---------?----------?---- 401
CONSENSUS-CPZ T.....TG--------.----------CT----------T-CAC-G 488
TAR CORE ->/
-LBP-1- UBP-2 ->/ signal ->
CONSENSUS-A GACCAGAT??GAGcCTGgGAgCTCTCTGGCTAGCGagGG.AACCCACT.GCTTAAGCCTCAATAAAGCTT 529
CONSENSUS-B --------ct----------T----.--------?---------- 535
CONSENSUS-D --------TT-------a----------T----.--------.---------- 527
CONSENSUS-O ------G-CT-----C--------C-----CT.-T--CTG-----G--.------CG---------- 459
CONSENSUS-U ------?--------C----------.-A--.--------.---------- 465
CONSENSUS-CPZ --------TT----------A.-C--T--A--.--------.---------- 555

HIV1 LTR CONSENSUS



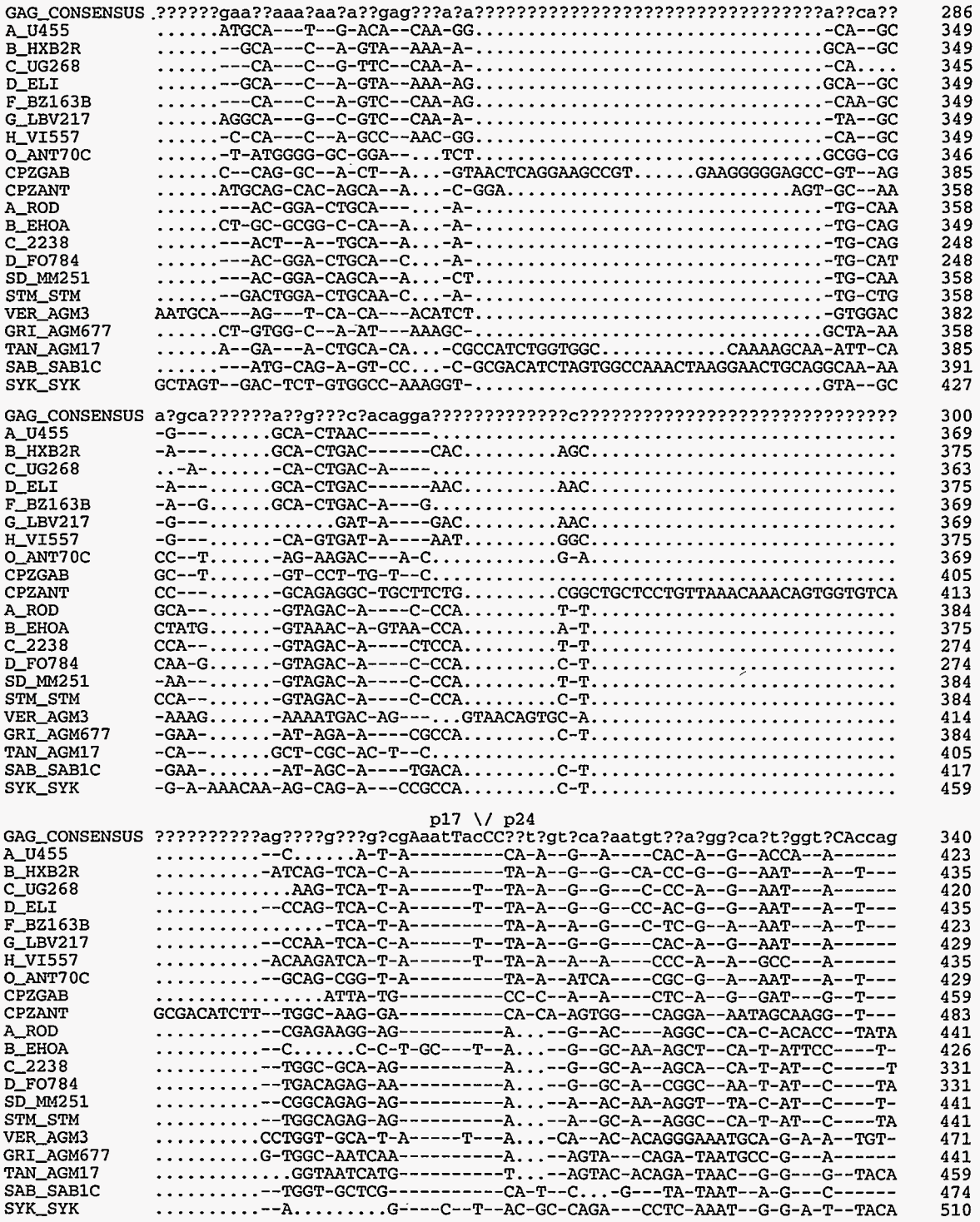
HMMER Sequences in the Gag Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_HXB2R	HIVHXB2R	K03455	Starcick,B.	Science 227, 538 (1985)
C_UG268	HIVUG268	L11799	Louwagie,J.J.	AIDS 7, 769 (1993)
D_ELI	HIVELI	K03454	Alizon,M.	Cell 46, 63 (1986)
F_BZ163B	HIV1BZ163B	S0585	Louwagie,J.J.	ARHR 10, 561 (1994)
G_LBV217	HIVLBV217	L11778	Louwagie,J.J.	AIDS 7, 769 (1993)
H_VI557	HIVVI557	L11793	Louwagie,J.J.	AIDS 7, 769 (1993)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68, 1586 (1994)
CPZANT	SIVCPZANT	U42720	Vanden Haesevelde,M.	Virology 221, 346 (1996)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
A_ROD	HIV2ROD	M15390	Clavel,F.	Nature 324, 691 (1986)
B_EHOA	HIV2EHOA	U27200	Rey-Cuille,M.A.	Virology 202, 471 (1994)
SD_MM251	SIVMM251	M19499	Franchini,G.	Nature 328, 539 (1994)
STM_STM	SIVSTM	M83293	Novembre,F.J.	Virology 186, 783 (1992)
VER_AGM3	SIVAGM3	M30931	Baier,M.	Virology 176, 216 (1990)
GRI_AGM677	SIVAGM677	M66437	Fomsgaard,A.	Virology 182, 397 (1991)
TAN_AGM17	SIVAGM17	L19250	Hirsch	Virology 197, 426 (1993)
SAB_SAB1C	SIVSAB1C	U04005	Jin,M.J.	EMBO J. 13, 2935 (1994)
SYK_SYK	SIVSYK	L06042	Hirsch,V.M.	JVI 67, 1517 (1993)

The following alignment was generated using the HMMER program. For simplicity, only representative types and subtypes are shown. An ordinary consensus sequence (lowercase signifies majority, uppercase signifies 50% or greater) was created from these sequences using MASE; this is not a "most likely sequence" based on a HMMER model (Part II). Annotation is based on HIV-1s, therefore the user shouod be cautious about its applicability to other PI sequences. Cleavage sites are designated by '\/'.

```
GAG_CONSENSUS ATGGGtGcgagagcgTCa???gtatTaagagGgagaaaattaGAt??aTggGAaaaaaTtaggtTacggC     65
A_U455        ------------------...--------C----A----------TC------G------C-----A---     67
B_HXB2R       ------------------...--------C---G--G--------CG----------C-----A---     67
C_UG268       ------------------...----------CG----------ACC----------A----A---     67
D_ELI         ------------------...--------C---G----------AA----------C---------     67
F_BZ163B      ------------------...--------C---G-----C-----GC----------C-----A---     67
G_LBV217      ------------------...--------C---G----------GA----------C-GA---     67
H_VI557       ------------------...--------C---G----------GC------G------C----GA---     67
O_ANT70C      ----------T-----T...--G--G-C---A--C----G---GC-------C----------AA--     67
CPZGAB        ------------------...--TC---C----G----------CGC---------G--C--C-TA---     67
CPZANT        -----A---G-G-----T...--T--G--G--AGAG--GC-----AC--------GT--C---C-T----     67
A_ROD         -----C------AAC--C...--C--G-------A----GC----GA--TA----G---C----------     67
B_EHOA        -----C-------GC--C...--C--GTC-----AG--GAC----GA--TA-----GG----------     67
SD_MM251      -----C------AAC--C...--C--GTC-----AG---GC----GA--TA----------C----A-     67
STM_STM       -----C------AGC--C...--C--GTC-----A----GC----GA--TA------G----------     67
VER_AGM3      -----G---GCTA-C---...-C-C---ATA-----C-------CAA--TT--GC-T--AC-AC-T--C-     67
GRI_AGM677    -----C-G-G-TCAC---...-C-C-GTC--------GCC-C--CACG-TC--G--G------C----T-     67
TAN_AGM17     -----A--AG--CAC--G...-C-C-GTC-A----G--T--G--CAC--TT--------AC-C-----C-     67
SAB_SAB1C     ----------TAAC---...--C-----T--T--G---C-----GC--TC---TC-G-C----------     67
SYK_SYK       -----A--AGCG-G---GGCAA------C---T---G-GC----CAG--AT--G------C--C-C--A-     70

                      <- INS-1 in HIV-1s
GAG_CONSENSUS C?gg?ggaAagAAaaa?TAta?a?T?aaaCAt?TagTaTgGgcAagcaaggAgcTgga?agaTttgca?t    126
A_U455        -A--G-----C-----A----G-C-G------T---------------G---------A-A---CA--C-    137
B_HXB2R       -A--G----------A----A-T-A------A---------------G------A--AC----C---G-    137
C_UG268       -A--G----------TGC----TGA-G-----CC----C---------G--------GA----------C-    137
D_ELI         -A--A----------A----G-C-A------A---------------G------A--AC---A----C-    137
F_BZ163B      -G--G----------A----GCT-G--G---C---------------G------A--AC--------A-    137
G_LBV217      -A--G----------A----G-A-G------T---------------G----------G---------C-    137
H_VI557       -A--G----------A----G-C-G------T---------------G----------A---------C-    137
O_ANT70C      -A--ATCT--A--G--A----G-C-A-----CT--------------GA---------A-----C---TG    137
CPZGAB        -C--G----GA----GA----TGA-G------T---------------G---------A-----C---TG    137
CPZANT        -C--T--C--------G--C-TGA-A------C-G--T--------ATC--------C-GC-T-----GC-    137
A_ROD         -C--C----------G--C-GGC-A------A-T--G------GCG--TA-AT----C-----C-G-T-    137
B_EHOA        -C--C----------GA--C-TGC-A------A----------GTG--T--AT-A--G-------G-T-    137
C_2238                                                 %%-GCG--C--AT----C---------T-     27
D_FO784                                                %%-GCA--T--AT----C-------G-T-     27
SD_MM251      -C--C----------G--C-TGT-G--G---G----------GCA--T--AT-A--T-------G-T-    137
STM_STM       -C--C--G--------G--C-TGT-G------G----G-----GCA--C--AT----C-------G-T-    137
VER_AGM3      -GACC----------G--CCA-A-T------T--A--------G-------AA----GC-C--C-GCC-    137
GRI_AGM677    -GAAC--G--A--G--G--CCA-A-T------T--A--------G-A--A--AA----AC------GGT-    137
TAN_AGM17     -GAAC----------G---CA-C-A------T--A---A----G--------A----GC-C----GCC-    137
SAB_SAB1C     -CAAC--------G--G----AGT-A-G----C-------------------A--A--TC-----AGCC-    137
SYK_SYK       -GAAA--G-----GCGC--CCT-G-TCG----T----T-------AG------T----C-----C-GCC-    140

GAG_CONSENSUS ??a?gc??gcct?tTaGaaacaaaaGAaGG?tgtcaaaAaat??t???agt?tTa?a?cCatt?gt?ccg    179
A_U455        TA-CC-TG----T---------GC------A-----GC----AC-GGG-CAA---C-A---GCTC-C-A-    207
B_HXB2R       TA-TC-TG----G---------TC------C---AG-C----AC-GGG-CAGC--C-A----CCC-T-A-    207
C_UG268       TA-CT-TG----T---------TC------C---A--C----AA-GAA-CAGC--C-A---GCTC-T-A-    207
D_ELI         TA-TC-TG----T---------TC------C---A--C----AA-AGGGCAGC--C-A---GCTA-T-A-    207
F_BZ163B      TA-TC-TG----TC--------TC------C----------AA-AGG-CAG---C-AT-G-CCC-T-A-    207
G_LBV217      TA-CC-TG----C----------C------T------C----AC-GCA-CAG--GC-A----CTC-C-A-    207
H_VI557       CA-CC-CGA---T-----C---GC------C----T-C--T-AA-AGA-CAGC--C-A---GCTC-TAA-    207
O_ANT70C      TA-TC-TGAG--AC--------TGC---G--TAA-G-G--GC-GC-ACAGCAG---G-G---GCTC-CAA-    207
CPZGAB        TG-CC-CG-G--AA-G----GT--G-----A---ACT---T-GT-ACA-CAA---G-G---GCTC-CAAA    207
CPZANT        CAGCT-CTC---TC--------TC------T---G----GGCTA-CCATCAA--GAGC--T-CCA-AGAA    207
A_ROD         AGCA-AGA----G--G--GT-------G--T----------TC-TAC---T---G-T---A-G--A---    207
B_EHOA        GGCA-AGA---GG--G-G-T----------A--C-GG-----TCGGAA---C---GGA-----A--A--A    207
C_2238        AGCA-AAA----G--G--G--C--------A--C--------TT-AAC---G---G-A-----A--T---     97
D_FO784       AGCA-AAA----G--G----AT--------G----------TC-CTCG--C---GCT-----A--A---     97
SD_MM251      AGCA-AAA----G--G--G-AC--------A----------AC-TTCG--C---GCT-----A--G--A    207
STM_STM       GGCA-AAA----G--G--GT----------G----------TA-AACG--C---G-A-----A--G--A    207
VER_AGM3      CC-T-AGA-AT-AC-----T--G-------A---A-G--G--CA-AGA---AC-CT-C--GC-A-AA--A    207
GRI_AGM677    AC-T-AGAAA--T----------------C----------CA-AGA---T---ACC--G--G-AA---     207
TAN_AGM17     CC-T-AGAAA--T--------------G--A----------CA-AGA---A---GCC--T--A-AA--A    207
SAB_SAB1C     TTCA--AAAT--C-----G-----G-----GGTCGTC-----TC-TAGT--AC-CCTT--CC-A--T--T    207
SYK_SYK       TTCG-ACCAA--CA-G--GT-T--------C---G-------TC-CAC---CC--CTT-----A-AAG-T    210
```

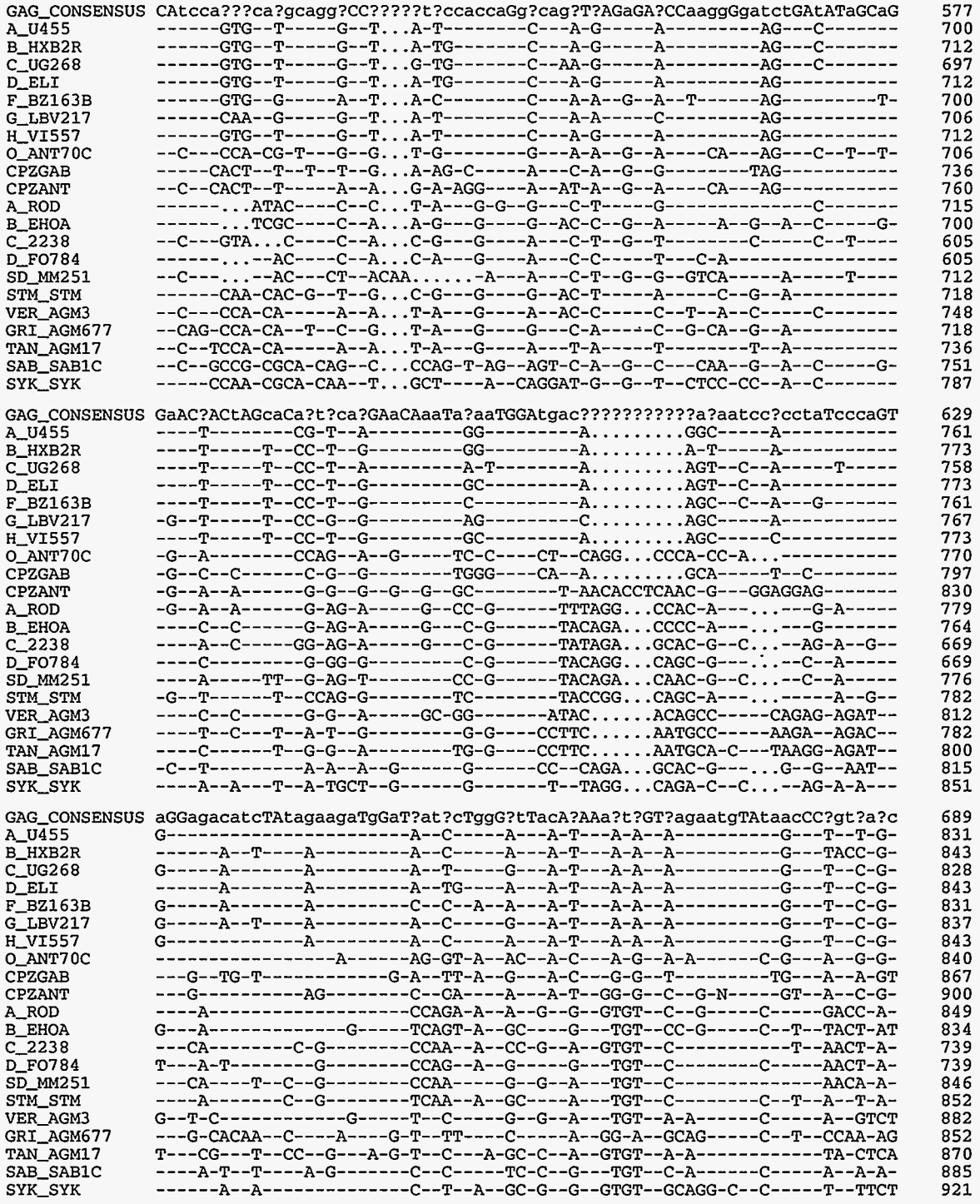
```
GAG_CONSENSUS AcagG?tCagaaga??Taaaa?c?cT?Tttaatac?gt?tg?gtc?Tctggtgt?TaCAc??agaa?a?a    234
A_U455        -----AA-------AC-T-G-T-AT-A-A------A--AGCA---C---AT---G----TCA-AGGATAG    277
B_HXB2R       -----A--------AC-T-G-T-AT-A-A------A--AGCAAC-C---AT---G-G--TCA-AGGATAG    277
C_UG268       -----GA----T--AC-T-G-T-AT-A-A---C--A--AGCAACTC---AT---G----TAA--GGATAG    277
D_ELI         -----AA-------AC-T-G-T-AT-A-A------A--AGCAAC-C---AT---G----TAA--G-ATAG    277
F_BZ163B      -----A--------GC-T-G-T-AT-A-A------A--AGCA---C---AT-T-G----TCA-AG-GTAG    277
G_LBV217      -----AA----G--GC-T---T-AT-A------G-A--AGCA--TC--------A----TCA-AGGATAG    277
H_VI557       -----AA-------AC-TC--T-AT-A------CTAT-AGCAAC-C---AT--CG----TCA-AG-ATAG    277
O_ANT70C      -----G-----CAGCC-GC-GT-A--C-GG--CG-AA-AGTA--GC--------G-T---AACAG-T-T-    277
CPZGAB        -----C-------GAC-GCGGT-CT-G-----C--TC-GGCA--AC-G-----CA----TAGT--CATC-    277
CPZANT        -T-A-A--CCCT--AA---T-T-TT-G-----C--CA-T--T--TC-G-----CG----TAA--GGG-A-    277
A_ROD         -----T------A-TT-----AGT--T--------T--C--C---A-T-----CA-----GC----G-G-    277
B_EHOA        -----T------A-TT-----AGT--T-A------C--C--C---A---TT--CC-G---GC----G-G-    277
C_2238        -----T------A-TT----GAGC--T-A------TACC--C---A-T------T-----GC----C-G-    167
D_FO784       -----T------A-TT-----AGC--T-AC-----T--C--C---A--------T-----GC----G-G-    167
SD_MM251      -----C------A-TT-----AGC--T-A------T--C--C---A-------CA-T---GC----G-G-    277
STM_STM       -----T------A-TT----GAGT--T--------T--C--C---A-T------A-C---GC----G-G-    277
VER_AGM3      -----G--G--G-GCT-----AGT--G-----CCTT--G--C--AT-G-TT--CG-----AA---TA-GG    277
GRI_AGM677    -----C--C--G-GGC-----G-T--G------TTGTGC--C---A-T-----CA-T---GC----C-G-    277
TAN_AGM17     -----A--G----GCT-G---AGT--G------TTGTGC--C---A-T------A-T---GC----C-G-    277
SAB_SAB1C     -----G-----GA-TT---TTG-T--C------TTATGC--C--GT-AGCT---A-T---GC----ATA-    277
SYK_SYK       -AT--G------A-TT-----AGC--T---GGC-TAA-T-CG---G-A---GCCG-T--TGCCA--A-GG    280

                        INS-1 ->
GAG_CONSENSUS aagT?aaagAcAC?gAggaagc??taga?aaa??????????????????tagag????????????ga    268
A_U455        -T--A--------CA-------TT--A-T---.................A-----............--    317
B_HXB2R       -GA-A--------CA-------TT----C--G.................A-----............--    317
C_UG268       ----ACG------CA-------CT----C--G.................A-----............--    317
D_ELI         -T--A--------CA-------TT----A--G.................A-G---............--    317
F_BZ163B      -G--A--G-----CA-------TT----C--G.................C-----............--    317
G_LBV217      GG--A--------CA-A-----TC----GG--.................G-G--A...........A-    317
H_VI557       -T--A--------CA-------TT----G--G.................T--A--............--    317
O_ANT70C      --A-TGG---T--GC--C-G--AA--C-A--G.................T--A--............--    317
CPZGAB        CT--AG-------AC--A----TC----AC-GCTAAAG......CGGCATCAT-GA............--    329
CPZANT        -GA-A--------A--AC----CG-TA-A-C-.....GTGAAAA-GA-A.....-T    323
A_ROD         ----G-----T--T--A-G---AAA-C-A-T-GTGCGG......AGACATC---T-.....-C    329
B_EHOA        ----G-----T--T--------AAA-....-G.................A---CACAGCGA...CAT..    321
C_2238        ----G---C-T--A------CGAAAC--GGT-GTAGAG......AGACATC---C-.....-C    219
D_FO784       ----G--------T----G---AAA-C-G-T-GTGCAG......AGACATC---T-.....-T    219
SD_MM251      ----G---C----T--------AAA-C-G-T-GTGCAG......AGACACC---T-.....-T    329
STM_STM       ----G---C-T--T--------AAAGC-AGTGGTAAAG......AGACATC-T-T-.....-T    329
VER_AGM3      ----G--------A--A-----AG---CA-T-GTAAGACAATGCTGCCATC---T-.....--    335
GRI_AGM677    ----G--------A--------TG---TA-C-.................G-TA--CAACACTACCAT..    327
TAN_AGM17     ----G--------A--------TG---TA-T-GTAAAACAACACTGCCATC---T-.....--    335
SAB_SAB1C     ----G-----T--A-------AAA--CG---GTAAAA......GAGGAAG-GCCA.....-C    329
SYK_SYK       ----AG-------T---C----AAA-C-G---.................G--A-A.....--    320

GAG_CONSENSUS a????????????????????????????????????????????????????????????????    269
A_U455        -.....................................................................    318
B_HXB2R       -.....................................................................    318
C_UG268       -.....................................................................    318
D_ELI         -.....................................................................    318
F_BZ163B      -.....................................................................    318
G_LBV217      G.....................................................................    318
H_VI557       G.....................................................................    318
O_ANT70C      -.....................................................................    318
CPZGAB        -.....................................................................    330
CPZANT        -.....................................................................    324
A_ROD         -.....................................................................    330
B_EHOA        .....................................................................    321
C_2238        -.....................................................................    220
D_FO784       G.....................................................................    220
SD_MM251      G.....................................................................    330
STM_STM       -.....................................................................    330
VER_AGM3      GAAAGAAAGA............................................................    345
GRI_AGM677    .....................................................................    327
TAN_AGM17     -.....................................................................    336
SAB_SAB1C     -.....................................................................    330
SYK_SYK       -GCATGTAATTGGAAAGATGACCCACCAGCGACATCTGGTGGACAAAGTGAAAATAGCAGTCAAAACATG    390
```



```
GAG_CONSENSUS .??????gaa??aaa?aa?a??gag???a?a?????????????????????????????????a??ca??   286
A_U455        ......ATGCA---T--G-ACA--CAA-GG.................................-CA--GC   349
B_HXB2R       .......--GCA---C--A-GTA--AAA-A-................................GCA--GC   349
C_UG268       .......---CA---C--G-TTC--CAA-A-.................................-CA....   345
D_ELI         .......--GCA---C--A-GTA--AAA-AG................................GCA--GC   349
F_BZ163B      .......---CA---C--A-GTC--CAA-AG.................................-CAA-GC   349
G_LBV217      ......AGGCA---G--C-GTC--CAA-A-..................................-TA--GC   349
H_VI557       ......-C-CA---C--A-GCC--AAC-GG..................................-CA--GC   349
O_ANT70C      ......-T-ATGGGG-GC-GGA--...TCT..................................GCGG-CG   346
CPZGAB        ......C--CAG-GC--A-CT--A...-GTAACTCAGGAAGCCGT.....GAAGGGGGAGCC-GT--AG   385
CPZANT        ......ATGCAG-CAC-AGCA--A...-C-GGA.............................AGT-GC--AA   358
A_ROD         .......---AC-GGA-CTGCA---...-A-..................................-TG-CAA   358
B_EHOA        ......CT-GC-GCGG-C-CA--A...-A-...................................-TG-CAG   349
C_2238        .......---ACT--A--TGCA--A...-A-..................................-TG-CAG   248
D_FO784       .......---AC-GGA-CTGCA--C...-A-..................................-TG-CAT   248
SD_MM251      .......---AC-GGA-CAGCA--A...-CT..................................-TG-CAA   358
STM_STM       .......--GACTGGA-CTGCAA-C...-A-..................................-TG-CTG   358
VER_AGM3      AATGCA---AG---T-CA-CA---ACATCT...................................-GTGGAC   382
GRI_AGM677    ......CT-GTGG-C--A-AT---AAAGC-...................................GCTA-AA   358
TAN_AGM17     ......A--GA---A-CTGCA-CA...-CGCCATCTGGTGGC.............CAAAAGCAA-ATT-CA   385
SAB_SAB1C     .......---ATG-CAG-A-GT-CC...-C-GCGACATCTAGTGGCCAAACTAAGGAACTGCAGGCAA-AA   391
SYK_SYK       GCTAGT--GAC-TCT-GTGGCC-AAAGGT-...................................GTA--GC   427

GAG_CONSENSUS a?gca??????a??g???c?acagga?????????????c?????????????????????????????   300
A_U455        -G---......GCA-CTAAC------.......................................... .   369
B_HXB2R       -A---......GCA-CTGAC------CAC.........AGC.......................... .   375
C_UG268       ..-A-.......-CA-CTGAC-A----.......................................... .   363
D_ELI         -A---......GCA-CTGAC------AAC.........AAC.......................... .   375
F_BZ163B      -A--G......GCA-CTGAC-A---G.......................................... .   369
G_LBV217      -G---............GAT-A----GAC.........AAC.......................... .   369
H_VI557       -G---.......-CA-GTGAT-A----AAT.........GGC.......................... .   375
O_ANT70C      CC--T......-AG-AAGAC---A-C............G-A.......................... .   369
CPZGAB        GC--T......-GT-CCT-TG-T--C.......................................... .   405
CPZANT        CC---......-GCAGAGGC-TGCTTCTG.........CGGCTGCTCCTGTTAAACAAACAGTGGTGTCA   413
A_ROD         GCA--......-GTAGAC-A----C-CCA.........T-T.......................... .   384
B_EHOA        CTATG......-GTAAAC-A-GTAA-CCA.........A-T.......................... .   375
C_2238        CCA--......-GTAGAC-A----CTCCA.........T-T.......................... .   274
D_FO784       CAA-G......-GTAGAC-A----C-CCA.........C-T.......................... .   274
SD_MM251      -AA--......-GTAGAC-A----C-CCA.........T-T.......................... .   384
STM_STM       CCA--......-GTAGAC-A----C-CCA.........C-T.......................... .   384
VER_AGM3      -AAAG......-AAAATGAC-AG---...GTAACAGTGC-A.......................... .   414
GRI_AGM677    -GAA-......-AT-AGA-A----CGCCA.........C-T.......................... .   384
TAN_AGM17     -CA--......GCT-CGC-AC-T--C........................................... .   405
SAB_SAB1C     -GAA-......-AT-AGC-A----TGACA.........C-T.......................... .   417
SYK_SYK       -G-A-AAACAA-AG-CAG-A---CCGCCA.........C-T.......................... .   459

                                                  p17 \/ p24
GAG_CONSENSUS ??????????ag????g???g?cgAaatTacCC??t?gt?ca?aatgt??a?gg?ca?t?ggt?CAccag   340
A_U455        ..........--C......A-T-A---------CA-A--G--A----CAC-A--G--ACCA--A------   423
B_HXB2R       ..........-ATCAG-TCA-C-A---------TA-A--G--G--CA-CC-G--G--AAT---A--T---   435
C_UG268       ............AAG-TCA-T-A------T--TA-A--G--G---C-CC-A--G--AAT---A------   420
D_ELI         ..........--CCAG-TCA-C-A------T--TA-A--G--G--CC-AC-G--G--AAT---A--T---   435
F_BZ163B      ..............-TCA-T-A---------TA-A--A--G---C-TC-G--A--AAT---A--T---   423
G_LBV217      ..........--CCAA-TCA-C-A------T--TA-A--G--G----CAC-A--G--AAT---A------   429
H_VI557       ..........-ACAAGATCA-T-A------T--TA-A--A--A----CCC-A--A--GCC---A------   435
O_ANT70C      ..........--GCAG-CGG-T-A---------TA-A--ATCA----CGC-G--A--AAT---A--T---   429
CPZGAB        ..............ATTA-TG----------CC-C--A--A----CTC-A--G--GAT---G--T---   459
CPZANT        GCGACATCTT--TGGC-AAG-GA----------CA-CA-AGTGG---CAGGA--AATAGCAAGG--T---   483
A_ROD         ..........--CGAGAAGG-AG----------A...--G--AC----AGGC--CA-C-ACACC--TATA   441
B_EHOA        ..........--C......C-C-T-GC---T--A...--G--GC-AA-AGCT--CA-T-ATTCC----T-   426
C_2238        ..........--TGGC-GCA-AG----------A...--G--GC-A--AGCA--CA-T-AT--C-----T   331
D_FO784       ..........--TGACAGAG-AA----------A...--G--GC-A--CGGC--AA-T-AT--C----TA   331
SD_MM251      ..........--CGGCAGAG-AG----------A...--A--AC-AA-AGGT--TA-C-AT--C----T-   441
STM_STM       ..........--TGGCAGAG-AG----------A...--A--GC-A--AGGC--CA-T-AT--C----TA   441
VER_AGM3      ..........CCTGGT-GCA-T-A-----T---A...-CA--AC-ACAGGGAAATGCA-G-A-A--TGT-   471
GRI_AGM677    ..........G-TGGC-AATCAA----------A...--AGTA---CAGA-TAATGCC-G---A------   441
TAN_AGM17     ............GGTAATCATG----------T...--AGTAC-ACAGA-TAAC--G-G---G--TACA   459
SAB_SAB1C     ..........--TGGT-GCTCG----------CA-T--C...-G---TA-TAAT--A-G---C------   474
SYK_SYK       ..........--A.........G----C--T--AC-GC-CAGA---CCTC-AAAT--G-G-A-T--TACA   510
```

```
GAG_CONSENSUS cc??Tatcacc?aGaAC?tTaAAtgCaTGGGTaaAagtagTagAagaaAAga??Ttcag?gCagaagTag   403
A_U455        G-CT-------T--G--C--G-----------G--------------C---GCT-----CC--------A   493
B_HXB2R       G-CA-------T-----T-----------------------------G---GCT-----CC--------GA  505
C_UG268       G-CC-------T-----T--G-------------------A----G------GCT-----CC----G---A   490
D_ELI         G-CA-------T-----T--G--C----------------A-----------GCT-----CC--------A   505
F_BZ163B      T-CA-------T-----T-----------------G--GA-------G---GCT--T--CC--------A   493
G_LBV217      --CA-------T--G--T-------------------------------GCC-----TC--------A     499
H_VI557       G-CA-------T--G--C-------------------------------G---GCT--T--CC--------A   505
O_ANT70C      G-CA-C--C--C--G--T-----------------G-C-----------AGCC--T-ACC-T---A-CA     499
CPZGAB        G-CA-C-----C-----C--G-----------------G-----G------GCA--T--CC-T--G---A     529
CPZANT        --AC-GA----A-----C--------C--------GTGT--------G--A-AT----ATC-------CA     553
A_ROD         --GC-GAGT--CC----CC--------C---------T-------G-----A-AG---G-G----------   511
B_EHOA        --GC--AGC--CC----C--------C---------T----------G----AG---G-G----------   496
C_2238        --TC--AGT--T-----AC-G--------------GT---------------AG--TG-G-----G----   401
D_FO784       --GC-GAGC--C-----A-------------T--GT----G--G--C--A-AA---G-G-----G----   401
SD_MM251      --AT--AGC--G-----A--------C---------T-GA----G-------AA--TG-A----------   511
STM_STM       --AT--AGC--A-----C--G-------------GT----G--G-------AA--TG-T--------G-   511
VER_AGM3      --CT-G-----AC-C--C--------G----------C------G--G--A-AA---G-A------A---   541
GRI_AGM677    --TT-G--T--GC-C--G--------G-----C---TGC--G--G-----A-GG-GGG-A----------   511
TAN_AGM17     --AC-G-----AC-C--CC-G-------------GAC---T--------A-GG--TG-A------A---   529
SAB_SAB1C     --AT-G--C--A--G--AC-T-----T-----------GA-T--G-------AG-----T-----G----   544
SYK_SYK       GGAG-CC--GTT--G--TC----AA-------GG-G-CT--GA-CTCC--A-AA--TGAT---TCCA---   580

GAG_CONSENSUS TaCCcatgTTt???GCacTatCaGAaGGagccac?cc?tAtGAt?T?AAt???aTGcTaAAt??agTgGG   461
A_U455        -----------TCA---T-------G--------C--AC-A---T-A---ATG-----G---GT------   563
B_HXB2R       -----------TCA---T----------------C--AC-A---T-A--CACC--------CAC------   575
C_UG268       -----------ACA---T----------------C--AC-A---T-A--CACC---------AC------   560
D_ELI         -----------TCA---T----------------C--AC-A---T-A--CACC--------CAC------   575
F_BZ163B      -----------TCA---T----------G-----T--AC-A---T-A---ACC---T-----AC------   563
G_LBV217      -----------TCA---T----------------T--AC-A---C-A---ACC---------AC------   569
H_VI557       -----------TCA---T----------------C--AC-A---T-A---GCC---T-----AT------   575
O_ANT70C      -T--T-----CATG---T-G-----G-----A-TTT-C------A-T---ACT---------GCCA-A--   569
CPZGAB        ----A------TCA--T--------G------TTA--TC-G---G-T--CACC---T-----GC---A--   599
CPZANT        -C--T------TCT--TT----------G--A--T--TC-----T-A--CACC-----T---GC---T--   623
A_ROD         -G--AGGA---CAG-----C--------CTG---G--C------A-C--CCAA-----T---TGT-----   581
B_EHOA        ----AGGA---CAG--------------GTG---C--T------A-T---CAG---T-----TGT--A--   566
C_2238        ----AGGA---CAG-----GA-------GTG---A--C------A-T---CAAT--------TGT--A--   471
D_FO784       -G--AGG----CAG--------G-----CTG---T--G-----CA-C---CAG---------TGT--A--   471
SD_MM251      -G--AGGA---CAG-----G--------CTG---C--C-----CA-T---CAG---T-----TGT-----   581
STM_STM       -G--AGGA---CAG--------------CTGT--C--C-----CA-C---CAAC--T-----TGT-----   581
VER_AGM3      -G--------CCAG--TT----------GTG---A--C-----CA-C---CAA-----T---GTCC----   611
GRI_AGM677    -C---------CCAA-----C-----G---TGTCTCT-C------G-A---CAG-----C---GT-A-A--   581
TAN_AGM17     -G---------CAG--T--C--------CTG-CTTAGC-----CA-T---CAA---T-----GTCA-A--   599
SAB_SAB1C     -----------AGC--C---G----------T-TC--T------A-C---CAA-----C---GC---A--   614
SYK_SYK       -G--AT-A---CAG--T---A--------TTT-TT--T------C-C---GGC-----T---GCT--A--   650

GAG_CONSENSUS ?GaaCAtCAaGcaGCcaTgCAaaT??TaAaaGA?at?ATcAATGAgGAaGc?Gcaga?TgGGAtagg??a   522
A_U455        G-G---C--G-----T--------GT-------T-CC--------------T-----G-----C---TT-   633
B_HXB2R       G-G----------------------GT-------G-CC--------------T-----A--------AGTG  645
C_UG268       G-G----------------------GT-------C-CC-----------G--T-----A---------TT-   630
D_ELI         G-G----------------------GC-------G-CC--------A-----T-----A---------TT-   645
F_BZ163B      G-G----------------------GC-------C-CC--------------T-----A-----C--ATT-   633
G_LBV217      G-G----------T---------GC----G--T-CT--------------T-----G---------AT-   639
H_VI557       G-G----------------------GT-------T-CA--------------T-----A---------GT-   645
O_ANT70C      A-G--------G---TT-A---G-GC----G--AG-A--------------A-T---G--------AACT   639
CPZGAB        G-G--------G----------G-GC----G--AG-C--T-----------C--C--G------C-CTT-   669
CPZANT        G--C--------G---------GG-GC-------AG-A--------------A--T--G---------TT-   693
A_ROD         C--C-----------------G--AA-C-GG--G--T--------------A-----A------GT-CA-   651
B_EHOA        A--------G--------------TA-T-GG--G--A--T-----------A-----C-----CCAACAG   636
C_2238        A--T-----G--T--T-----G--TA-T-G---A--T--------------A--C--T------GAACAC   541
D_FO784       A--------G----------G--CA---GG--A--A--------T--G--A-----T------CA-CAG   541
SD_MM251      A--C--------G--T-----G--TA-C-G---T--T--A--------G--T-----T-----CTT-CAG   651
STM_STM       G--G----------------TA-------G--C--A----------T--T--T-----C-T-CAG      651
VER_AGM3      A--C-------GG--GC-A-----AG-------A--C--------------A--CC-G-------TAGCT   681
GRI_AGM677    A--C-----G-GG--AT-A-----TC-T--G--AG-C--T-----A-----A-----G-----C---AC-   651
TAN_AGM17     A--C--C-----G---G--------TA-T-----GG-A--T-----A--G--A--CC-G-A---CCTAACC  669
SAB_SAB1C     A-----C--G-GG--AC-A-----AG-G-----TG-T--------A-----G-----T-----CCTTAG-   684
SYK_SYK       A--T-----G-G---------GG-AA-T-----C--C--------------GG--G-CA------CTTGA-   720
```

HIV1 GAG


GAG_CONSENSUS CAtcca???ca?gcagg?CC?????t?ccaccaGg?cag?T?AGaGA?CCaaggGgatctGAtATaGCaG 577
A_U455 ------GTG--T-----G--T...A-T--------C---A-G-----A---------AG---C------- 700
B_HXB2R ------GTG--T-----G--T...A-TG-------C---A-G-----A---------AG---C------- 712
C_UG268 ------GTG--T-----G--T...G-TG-------C--AA-G-----A---------AG---C------- 697
D_ELI ------GTG--T-----G--T...A-TG-------C---A-G-----A---------AG---------- 712
F_BZ163B ------GTG--G-----A--T...A-C--------C---A-A--G--A--T------AG---------T- 700
G_LBV217 ------CAA--G-----G--T...A-T--------C---A-A-----C---------AG---------- 706
H_VI557 ------GTG--T-----G--T...A-T--------C---A-G-----A---------AG---------- 712
O_ANT70C --C---CCA-CG-T---G--G...T-G--------G---A-A--G--A----CA---AG---C--T--T- 706
CPZGAB -----CACT--T--T--T--G...A-AG-C-----A---C-A--G--G--------TAG---------- 736
CPZANT --C--CACT--T-----A--A...G-A-AGG----A--AT-A--G--A----CA---AG---------- 760
A_ROD ------...ATAC----C--C...T-A---G-G--G---C-T-----G--------------C------- 715
B_EHOA ------...TCGC----C--A...A-G---G----G--AC-C--G--A-----A--G--A--C-----G- 700
C_2238 --C---GTA...C----C--A...C-G---G----A---C-T--G--T--------C-----C--T---- 605
D_FO784 ------...--AC----C--A...C-A---G----A---C-C-----T---C-A--------------- 605
SD_MM251 --C---...--AC---CT--ACAA.....-A---A---C-T--G--G--GTCA-----A-----T---- 712
STM_STM ------CAA-CAC-G--T--G...C-G---G----G--AC-T-----A-----C--G--A---------- 718
VER_AGM3 --C---CCA-CA-----A--A...T-A---G----A--AC-C-----C--T--A--C-----C------- 748
GRI_AGM677 --CAG-CCA-CA--T--C--G...T-A---G----G---C-A-----C--G-CA--G--A---------- 718
TAN_AGM17 --C--TCCA-CA-----A--A...T-A---G----A---T-A-----T--------T--A---------- 736
SAB_SAB1C --C--GCCG-CGCA-CAG--C...CCAG-T-AG--AGT-C-A--G--C---CAA--G--A--C-----G- 751
SYK_SYK -----CCAA-CGCA-CAA--T...GCT----A--CAGGAT-G--G--T--CTCC-CC--A--C------- 787
GAG_CONSENSUS GaAC?ACtAGcaCa?t?ca?GAaCAaaTa?aaTGGAtgac??????????a?aatcc?cctaTcccaGT 629
A_U455 ----T--------CG-T--A---------GG---------A.....GGC-----A---------- 761
B_HXB2R ----T-----T--CC-T--G---------GG---------A.....A-T-----A---------- 773
C_UG268 ----T-----T--CC-T--A---------A-T--------A.....AGT--C--A-----T----- 758
D_ELI ----T-----T--CC-T--G---------GC---------A.....AGT--C--A---------- 773
F_BZ163B ----T-----T--CC-T--G---------C----------A.....AGC--C--A---G------- 761
G_LBV217 -G--T-----T--CC-G--G---------AG---------C.....AGC-----A---------- 767
H_VI557 ----T-----T--CC-T--G---------GC---------A.....AGC-----C---------- 773
O_ANT70C -G--A--------CCAG--A--G-----TC-C----CT--CAGG...CCCA-CC-A...------- 770
CPZGAB -G--C--C------C-G--G--------TGGG----CA--A.....GCA-----T--C-------- 797
CPZANT -G--A--A------G-G--G--G--G--GC--------T-AACACCTCAAC-G---GGAGGAG------- 830
A_ROD -G--A--A------G-AG-A-----G--CC-G------TTTAGG...CCAC-A---...---G-A----- 779
B_EHOA ----C--C------G-AG-A-----G---C-G------TACAGA...CCCC-A---...---G------- 764
C_2238 ----A--C-----GG-AG-A-----G---C-G------TATAGA...GCAC-G--C...--AG-A--G-- 669
D_FO784 ----C---------G-GG-G---------C-G------TACAGG...CAGC-G---...--C--A----- 669
SD_MM251 ----A-----TT--G-AG-T--------CC-G------TACAGA...CAAC-G--C...--C--A----- 776
STM_STM -G--T------T--CCAG-G--------TC--------TACCGG...CAGC-A---...-----A--G-- 782
VER_AGM3 ----C--C------G-G--A-----GC-GG-------ATAC......ACAGCC-----CAGAG-AGAT-- 812
GRI_AGM677 ----T--C---T--A-T--G---------G-G----CCTTC......AATGCC-----AAGA--AGAC-- 782
TAN_AGM17 ----C-------T--G-G--A--------TG-G----CCTTC......AATGCA-C---TAAGG-AGAT-- 800
SAB_SAB1C -C--T---------A-A--A--G------G------CC--CAGA...GCAC-G---...G--G--AAT-- 815
SYK_SYK ----A--A---T--A-TGCT--G------G-------T--TAGG...CAGA-C--C...--AG-A-A--- 851
GAG_CONSENSUS aGGagacatcTAtagaagaTgGaT?at?cTggG?tTacA?AAa?t?GT?agaatgTAtaacCC?gt?a?c 689
A_U455 G----------A--C-----A---A-T---A-A--A----------G---T--T-G- 831
B_HXB2R ------A--T----A---------A--C-----A---A-T---A-A--A----------G---TACC-G- 843
C_UG268 G-----A-------A---------A--T-----G---A-T---A-A--A----------G---T--C-G- 828
D_ELI ------A-------A---------A--TG----A---A-T---A-A--A----------G---T--C-G- 843
F_BZ163B G-----A-------A---------C--C--A--A---A-T---A-A--A----------G---T--C-G- 831
G_LBV217 G-----A--T----A---------A--C-----G---A-T---A-A--A----------G---T--C-G- 837
H_VI557 G----------A---------A--C-----A---A-T---A-A--A----------G---T--C-G- 843
O_ANT70C ----------A------AG-GT-A--AC--A-C---A-G--A-A------C-G---A--G-G- 840
CPZGAB ---G--TG-T----------G-A--TT-A--G---A-C---G-G--T---------TG---A--A-GT 867
CPZANT ---G----------AG--------C--CA----A---A-T--GG-G--C--G-N-----GT--A--C-G- 900
A_ROD ----A----------CCAGA-A--A--G--G--GTGT--C--G-----C-----GACC-A- 849
B_EHOA G---A----------G-----TCAGT-A--GC----G---TGT--CC-G-----C--T--TACT-AT 834
C_2238 ---CA--------C-G--------CCAA--A--CC-G--A--GTGT--C----------T--AACT-A- 739
D_FO784 T---A-T--------G--------CCAG--A--G-----G---TGT--C---------C-----AACT-A- 739
SD_MM251 ---CA----T--C--G--------CCAA-----G--G--A---TGT--C----------AACA-A- 846
STM_STM ----A-------C--G--------TCAA--A--GC----A---TGT--C---------C--T--A--T-A- 852
VER_AGM3 G--T-C----------G-----T--C-----G--G--A---TGT--A-A------C-----A--GTCT 882
GRI_AGM677 ---G-CACAA--C----A----G-T--TT----C-----A--GG-A--GCAG-----C--T--CCAA-AG 852
TAN_AGM17 T---CG---T--CC--G---A-G-T--C---A-GC-C--A--GTGT--A-A----------TA-CTCA 870
SAB_SAB1C ----A-T--T----A-G-------C--C-----TC-C--G---TGT--C-A------C-----A--A-A- 885
SYK_SYK ------A--A----------C--T--A--GC-G--G--GTGT--GCAGG-C--C-----T--TTCT 921

```
GAG_CONSENSUS aTttTaGAcaTaaaaCAaGGgcCaAAaGaacc?TTcaaagacTAtGTaGAtaGgTTCTataAaac?tTaa   757
A_U455        -----G-------G-----------------C----GG--T----------A----T------TC-C-   901
B_HXB2R       ---C-G-------G------A-----G-----C--T-G-----------CC-------------TC---   913
C_UG268       -----G--------------------G-----C--T-G-------------C------T------C----   898
D_ELI         -----G-------G---G--A-----G-----T--T-G------------CC-------------TC---   913
F_BZ163B      -----G-------G------------G--G--C--T-G-----------C-------T------CC---   901
G_LBV217      -----G-------G-----------------C--T-G---T----------------T------T--G-   907
H_VI557       -----G-------G-----------------C--T-G--------------------T----G-T----   913
O_ANT70C      --C-----T--T--G--G--A-----G-----A--T-G-------------C-------C-----A----   910
CPZGAB        --CC-T-----C-G------------G--G--C--T-G---T-----G--C--A-----C-----T----   937
CPZANT        ---C----G----------A--------G--C----G---T-----G-----A----------AA-T-   970
A_ROD         --CC-------------G--A-----G--G--G---C--AG------------A-----C----GC--G-   919
B_EHOA        --AC-G--------G-----------------C---C--AG------------A-----C----GC---C   904
C_2238        ---C-----G-G-----G--T----------A--TC-GAGT                              781
D_FO784       ---C-G--T--------------------GA-G---C-GAG-                              781
SD_MM251      ---C----TG-------------------G--A--TC-GAG---------C--------C----GC----   916
STM_STM       ---------G----G---------------A-G---C--AGT--------C--------C--G-GT--G-   922
VER_AGM3      G-C----------G---G-----C------G-A-----------C---------------C---G-AA---   952
GRI_AGM677    G-CC-------TCG---G--A--T--------C---C-G----------C--A---------G-CC-G-   922
TAN_AGM17     G-------T--C-G------AG-C----G---A-------------------------C--G-CC-T-   940
SAB_SAB1C     ----------T----------T--------T-----G--T----------A-----C--GG-TC-G-   955
SYK_SYK       --C--------------G--C---------ATT-----------------C--A-----CC-TTGC----   991

GAG_CONSENSUS Gagc?GAgCAagCt?cacagga?GTaAAaaatTgGaTgACagaaaC?tTgtTggTcCAaAATGC?AAcCC   822
A_U455        ----T---------A-------T--------C--------------C---C------------G--T--   971
B_HXB2R       ----C----------T-------G-----------------------C----------------G-----   983
C_UG268       ----T--A------A----A--T---------------------C--C----------------G-----   968
D_ELI         ----C---------T-------T---------------------C----------------A-----   983
F_BZ163B      ----T---------A-------A-----GGG------------C--C----------------G-----   971
G_LBV217      ----T---------A-------T--------C-----------C--T----------------G-----   977
H_VI557       ----T--A------A-------G--G---GG------------C--C----------------G--T--   983
O_ANT70C      ----T---------A-T--A--A-----------------------TC--C-T--T--------C--T--   980
CPZGAB        -G--T---------T-------A--G--------------T--C--TC-C--A--G--G-----A-----  1007
CPZANT        ----A--A--G---T-----CCT--G---GCC--------------C-----AA----------C--T--  1040
A_ROD         -G--A--A---A-AGAT-CA-CA--G--G-----------CC----AC--C-A--A--------C-----   989
B_EHOA        -G--A------A-AGAC-CA-CA-----------------C----AC--C--A-T--G-----T-----   974
SD_MM251      ----A--A---A-AGATGCA-CA-----G-----------TC----AC--C--A-T--------T-----   986
STM_STM       ----A--A-----AGAT-CATCA--G--------------TCGG--AC--C-AA-T--------T--T--   992
VER_AGM3      ----T-----G--CT--GGA--A-------C-G------------T-A--AC-CA-T--G-----T--T--  1022
GRI_AGM677    ----A--A-----AC-------T--T---------------C----T---C-TA----G-----C--T--   992
TAN_AGM17     -GCGC--A---A--C-C-----C----------A------------TC--C-CA----G-----T--T--  1010
SAB_SAB1C     ----A-----GA-AGAC-CT-CT-----G--C---------C-GT-TC-----A----G-----C-----  1025
SYK_SYK       ----A-----G--GGAC-CATCA--C---GGG---C-A--TC----AC-----A-T--------C-----  1061

GAG_CONSENSUS aGAttgtaag???aT?tTaAaagcatTggga????a?gCtaC?tTaGAaGAaATG?TgacaGC?TGtCAg   880
A_U455        ---C------TCC--T----G------A--GCCAGGG-----A------------A-------A--C---  1041
B_HXB2R       ----------ACT--T--------------CCAGCG-----AC-----------A-------A------  1053
C_UG268       ----------ACC--T----G------A--GCCAGGA---T-A-----------A-------A------  1038
D_ELI         ------C---ACT--C--------------CCAC-G-----AC-----------A-------A------  1053
F_BZ163B      -------C--ACC--T--------T------CCAGGG-----A                            1014
G_LBV217      ----------ACC--C-----------A---CCAGGA-----AC-----------A-------A--C---  1047
H_VI557       ----------ACT--T----G------A---CAAGGG---T-AA----------A-------C--C---  1053
O_ANT70C      ------C--ACAG--T--G---T----A--GCCAGGA-----C--------G---A--GT---C------  1050
CPZGAB        ----------CAA--TC-------------CCTGGG--C--CC----G------A-------T-----A  1077
CPZANT        ------C--ACAC--CC-G--G--T------ACAGGA--CT-C------------T-A-----T-----A  1110
A_ROD         ---C-----ATTAG-GC------G-C-A--GATGA-CC----C--------G---C----C--C-----  1059
B_EHOA        ------C--ATTAG-GC-T--G-GC------ATGA-CC-C--A----------C-A-----C--C----  1044
SD_MM251      ------C---CTAG-GC-G--G-GGC----TGTGA-TC-C--CC----------C-----G--T-----A  1056
STM_STM       ---C--C--ATTGG-A--G----GTC----CATGA-TC-C--CC-------G---T-A-----A-----A  1062
VER_AGM3      ---C-----AGTC--CC----G-GCC-----ATGC-TC-C--TC----------T-A--T--C-----A  1092
GRI_AGM677    G--------ATTG--TC-G----G-------ATGA-TC-A--C--G--G------C-A-T---T--C---  1062
TAN_AGM17     ----CAC--ATTAG-A--G---AG-C-A--GATGC-TC-A--A----------C-A-----A-----A  1080
SAB_SAB1C     ----------ACAG-GC-C----GG--A--CATGA-TC-A--CC-------G---C-T--T--A--C--A  1095
SYK_SYK       ---A--C-GACAG--CC-C--G---A--...GTCA-AC----C--G--G------C--CA---T-----A  1128
```

HIV1 GAG

```
                                                    p24 \/
GAG_CONSENSUS GGagTaGGaGGacCagg?cAtAA?GCaagagT?aTgGc?GAagc?aTga?a?a?gc????caa?at????   933
A_U455        -----G--------C--C-----A-----G--TT----T--G--A----GTC-A-TA...---C-G....  1104
B_HXB2R       --------------C--C-----G--------TT----T-----A----GCC-A-TA...AC-A--TCA.  1119
C_UG268       -----G--------T--C--C--A--------GT----T--G--A----GCC-A--A...A-CA--....  1101
D_ELI         -----G--G--G--CA-C-----A--------TC----T--G--A----GCC-A--A...AC-A--TCAG  1120
G_LBV217      -----G--------CA-C-----A--------TT-A--T-----G----GCC-G--A...TC-GG-ACAG  1114
H_VI557       -----G-----G--T--C-----A----A---TT----T--G--A----GCC-A-TAAATAC-A--....  1119
O_ANT70C      ----------ACT-----G--C-----GC-A--A-----A---GCTACA--C...--GC-AGATT       1117
CPZGAB        ----------TCT-----G-----G--TC-A--T-----C---TC-ATG-TT...--GA--CAAG       1144
CPZANT        ----------C----CC-----G--------GT----A-----T---GCTTCT--T...A-TA--GCAC   1177
A_ROD         --G-----T--G-----C--G--A--T---T-A-----A--G--CC---A-G-G-TCATAGG-CC-....  1125
B_EHOA        --GA-------C-----G--G--G-----GC-A-----T--G--TT-A-A-G-A--CTTGAC-CC-....  1110
SD_MM251      --------G--------A--G--G--T---T-A-----A-----CC---A-G-G--CCTCGC-CCA....  1122
STM_STM       --G--T--G--------A--G--A--T---C-G-----A-----CT---A-G-G-TCTTC---CCA....  1128
VER_AGM3      -----G---------A-TT-C--A----A---G-----A---ATG---CA-A-TATG...---AGC....  1155
GRI_AGM677    ----------G---CAA-----G--T-AGC-A----TA---ATG----GTA-T-GA...--GA--....  1125
TAN_AGM17     ---A----G--------A--C--A--C-AGC-T-----G-----C---CA-C-AATG...---GGA....  1143
SAB_SAB1C     --CA-T------G--CAA-----G---C-TC-A-----G-----T----C-GCG--CTTT--GC-ACAAA  1165
SYK_SYK       -----C--------CTTA--C--G----A-C-C-----A--G--C---GT-ATG--C...--GC-G....  1191

                                                    \/ p7
GAG_CONSENSUS ????????????????c??c?at?atgaTgcag?????????????????a?aggc???ca???????a?   954
A_U455        ..............A-AAGC--A---------.................-G----...A-T...TTT-G  1136
B_HXB2R       ..............G-TA-C--A---------.................-G----...A-T...TTT-G  1151
C_UG268       ..............ATAAAC--A---------.................-G-A--...A-T...TTT-A  1133
D_ELI         TT............A-TA-AGCA--------.................-G----...A-T...TTT-A  1154
G_LBV217      CA............A-AG-C--A---------.................-A----...A-T...TTT-A  1148
H_VI557       ..............G-AGTT--G---------.................-A---T...A-C...TTC-A  1151
O_ANT70C      TGAAAGGAGGATACA-AG-AG-AT-C-----A.................-G---G...--A...AATCC  1163
CPZGAB        GG.........AGAG-AGATG-TT-CT-C---.................-AG--A...--A...GGTGC  1181
CPZANT        AG.........GGAA-CG-AG-CT-TC-----.................-G----...A-TGGAAAT-G  1217
A_ROD         ..............G-CC-T--CCCAT-CGCA.................GC--C-...--G........  1152
B_EHOA        ..............T-CA-C-ATCC-T-TGCC.................GCC-CT...--A...CCA-G  1142
SD_MM251      ..............GTGC-A--CCCTT-TGCA.................GC--C-...--G...AAG-G  1154
STM_STM       ..............GACC-AC-GCCCT-CGCA.................GC--C-...--A...CAGCA  1160
VER_AGM3      ..................CAG-AC--------A.................CAG---...GGT...CAG-G  1184
GRI_AGM677    .........................---G-C--A.................GTG--ACCT--G...AAA-A  1151
TAN_AGM17     ..................GTC-AC-----A---GGCCACTCAAAGGGAGGA-GG--A...AGA...GGA-G  1190
SAB_SAB1C     CA.........GTTGGGAAC--CT-TG-----CAAGGGGCA...AGACCC-GG--T...-CC...TTAGG  1217
SYK_SYK       ..................T-AG-A-AT---GTA.................CA---G...-CA...TCA-A  1220

GAG_CONSENSUS ggg?c?aagaaga????????????????attaaaTGtTtcaAtTGtGG?aAagaaGG?CAcatagcaAga  1005
A_U455        ---C-CG------.............-----G--------C-----C--------A---C----C-A-  1191
B_HXB2R       -AAC-A-----AGATT..........G----G----------------C--------G----C---C---  1209
C_UG268       ---C-CT-AG---ATT..........-------------------C--G-----G---C----C---  1191
D_ELI         ---C-C-----A-ATT..........-----G---------------C--------G----------A-  1212
G_LBV217      ---C-CG----A-AAT..........-----G------G-C-----C--G-----A--TC----C---  1206
H_VI557       ---C-A-------ATT..........-------------------C--------A---C----C---  1209
O_ANT70C      AATTAGG-A-G--ACT..........--A-----------C-----A-----G--A--T---------  1221
CPZGAB        ---C-C--A----AAA..........--A-----C--T--------G--------T---C----T---  1239
CPZANT        A--AGG--A----CCT..........C-C--------T--C--C--T-----G--C--T-CT------  1275
A_ROD         ....-AG----AGGCA..........T-------C-GG--C-----A--G-----G---TCG------  1206
B_EHOA        A-CAGGG-AG--GACA..........G-G-C---C-GG--C--C--C--G-CG--A--T-C---C--G  1200
SD_MM251      ---A-C-----AGCCA..........-----G----GG--------G--G--G--A---TCT-----G  1212
STM_STM       ---A-G---G-C-.............G-A-----C-GG--------G----C---G----CT----A-  1215
VER_AGM3      A--AAG-CC----CCCCCA.......G-A--G----A---------A---TTT--C--T--GCA----  1245
GRI_AGM677    ---C-CCC--G-GCCG..........C-A-----C--T--------C---TTT--A--T--GCA---G  1209
TAN_AGM17     A--A-C-CCTC-C.....................-----T--A--C--AC-GATT--A--TG-TCA--A-  1239
SAB_SAB1C     ---AAG-G-T---CCTCTGAACCCAAAT----------A---------A--GCCT--T---C-G--T---  1287
SYK_SYK       A--GAG-TC--TG.............--------C-A---------GC--ATT--G-----GCA--A-  1275
```

```
GAG_CONSENSUS aatTGcagggcccctag?aaaaagggcTGctggaaaTGtGGaaaggaaggaCAccaaatgaaagacTGca  1074
A_U455        -----T----------G----------T--------C--G--A-------------------------  1261
B_HXB2R       ----------------G----------T-----------------------------------T--T-  1279
C_UG268       ----------------G-----A-----T-----------G-----------------------T-  1261
D_ELI         ----------------G----------T----G---------------------C-A-----T----  1282
G_LBV217      ----------------G----G------T----------------G-----T----------------  1276
H_VI557       ----------------G--------A--T--------C-------------T--GC------------  1279
O_ANT70C      -----TC-A--A-----G--------T--------------C-------T-------------T----  1291
CPZGAB        -----T-AA--A--A--A-G---A-----T---CGG-----GC-A--G---------------------  1309
CPZANT        -------A------A--A-GG--A----------G-------C-------------GC-T---A----TC  1345
A_ROD         C-A---C-A--A-----A-GGC-------------G-----T---CC--------ATC----C-A-----C  1276
B_EHOA        C-G----A---------A-GGC----A--------------C--C--------ATC---TC-A-A---C  1270
SD_MM251      C-A-----A-----A--A-G-C----A----------------AATG-AC--TGTT---GCCA-A---C  1282
STM_STM       C-G--T-AA-G---G--A-G-C-A--T--T------------C-AC-G--C--T--G---GCCA-A---C  1285
VER_AGM3      C-A---CCT-AA--A--A--G-T-A-A----T---------G--ACC---G--TTT-GCA-----T----  1315
GRI_AGM677    G-A----A---A--A--AC-G-TCAAA----TT--G--C--C--AATT--C--TATGGCA----------  1279
TAN_AGM17     G----TCC-CGAGGGG-TCC--TTAAG---CTC--------C--ACC-A-G---ATGGCA--------T-  1309
SAB_SAB1C     TT---T-A---------G-G-C-A----------------C-GCCC--AC--T--G--------T--TC  1357
SYK_SYK       G-C----AAAAG--ACTT---GCTAAA----TT--T--------AAC--------T-GCA-GG-C---T-  1345

                                        pol orf start ->
                                                    p7 \/
GAG_CONSENSUS ca??????????????????????????ga?agacAgGc?aaTTTTTAgGgaaaggctgg?c??cc??  1110
A_U455        -T.........................--G--------T--------------ATT---C-TT--AA  1304
B_HXB2R       -T.........................--G--------T--------------GAT----C-TT--TA  1322
C_UG268       -T.........................--G--------C--------------GATT---C-TC--AA  1304
D_ELI         -T.........................--G--------T-------------G-ATT---C-TT--CA  1325
G_LBV217      --.........................--G--------T--------------AT----C-TT--CA  1319
H_VI557       --.........................--A--------T-------------ATT---T-TC--AA  1322
O_ANT70C      G-AAT.....................-GA-A------A----------C---TA----C-TC-GGG  1337
CPZGAB        --.......................-GA-------TG----------------T---C-TT--CG  1352
CPZANT        --GCAACA.................AATACAG--A-A-TA--------------CCGACCC-CA-GTG  1397
A_ROD         --.........................--T--------AGG---------ACTG---CCTTGGGGA..  1317
B_EHOA        --.........................--A--------GGG---------TTC---CCCTGGGGA..  1311
SD_MM251      --.........................--C--------GGG---------CCTT--TCCATGGGGA..  1323
STM_STM       --.........................--A-------TGGG---------CTTT---CCATGGGGA..  1326
VER_AGM3      G-GGA.....................----TA----------T-T---C--TGGATGGG  1355
GRI_AGM677    AG.....................A-TG-------A----------T-T---CATTGGGGAGG  1322
TAN_AGM17     G-.....................-GAG-------C---------A---G-ATGCCCG-AC--TA  1352
SAB_SAB1C     A-AAA.....................----TT----------TTT---CCCTGGGG-AG  1397
SYK_SYK       GGCAACCCAAGAGAAATCAAGGCCCTCCA-TTGC------T----------------A---GGGGT-TC  1415

GAG_CONSENSUS ????aagaggaggCCagg?AAtTttct????????????????????????????????????????  1132
A_U455        C...---G----------G-------CT.....................................  1329
B_HXB2R       C...---G-A--------G--------T.....................................  1347
C_UG268       C...---G----------G-----C--T.....................................  1329
D_ELI         C...---G-A-----G--G--C-----C.....................................  1350
G_LBV217      C...---G----------G--------T.....................................  1344
H_VI557       C...--AG----------G--------C.....................................  1347
O_ANT70C      G...GGC-C----------C----A-G-G.....................................  1362
CPZGAB        C...-GCG----------C--C---G-G.....................................  1377
CPZANT        GTGGGG-T-C--A-----G--C---G-G.....................................  1425
A_ROD         ...-----A-...--CC-C--C--C-CCGTGGCC................................  1344
B_EHOA        ...-----A-...--TC-C-----C-CCGTCCAGGCACCCCAGGGGATAGTGCCATCTGCGCCCCCGATG  1374
SD_MM251      ...-----A-...--CC-C-----C-CCATGGCT................................  1350
STM_STM       ...----A-...--CCAC--C--C-CCATGGCC.................................  1353
VER_AGM3      A...GC--AA...--CA-A-------CC.....................................  1377
GRI_AGM677    A...GC--AA...---A-A------G-G.....................................  1344
TAN_AGM17     A...GGTTCC-A---GC-C------T-A.....................................  1377
SAB_SAB1C     A...GG--AA...--CA-G-----C-CT.....................................  1419
SYK_SYK       C...G----CCC----CC--C--C-CA.....................................  1440
```

HIV1 GAG

```
GAG_CONSENSUS ??????????????????????caga?cagacca???????????????gagCCaaCaGCcCCaCCag?a?  1163
A_U455        ......................----G-------.....------------------------C-.  1362
B_HXB2R       ......................----G-------.....--------------------A-.  1380
C_UG268       ......................----G-------.....--A-----------------C-G  1363
D_ELI         ......................--A-G-------.....------------------------C-.  1383
G_LBV217      ......................----A-------.....------T----------C-C-.  1377
H_VI557       ......................----G---G---.....------------------------C-.  1380
O_ANT70C      ......................----GACC-G--.....C-C---T----------GATG.  1395
CPZGAB        ......................----A----A--.....--A-----G--------GAT-.  1410
CPZANT        ..............CAGAAGG--GAAGT-GTG.....----------T-----CAT-.  1464
A_ROD         ..............CAAGTT-C-CAGG-G-TG.....ACA--------A--C----TG.  1383
B_EHOA        AACCCAGCATTCGGCATGACA-CTCAGG-TG-G.....ATT---T-T--A--C--T-C-.  1428
SD_MM251      ..............CAAGTG--TCAGG-G-TG.....AC------T--T--C----AG.  1389
STM_STM       ..............CAAATA-CTCAGG-G-TG.....AC------T-----T----A-.  1392
VER_AGM3      ..............GCCGCTACTCTTG-GGTG.....--------T--G--C--TCC-C  1417
GRI_AGM677    ....................--ATA----GG-GACACAGTTGGTCTG--A----------C---ATGG  1393
TAN_AGM17     ..................GAG--AGGAG--G-C.....-TT--G--G--T--C--GATCG  1414
SAB_SAB1C     ....................-T--C-TCCATC.....AGA--------T--C---ATGG  1453
SYK_SYK       ..........................GTG--GAGC.....------T-T--T--C--TCTGG  1471

GAG_CONSENSUS ????????????????????????????????????????????????????????????  1163
A_U455        ............................................................  1362
B_HXB2R       ............................................................  1380
C_UG268       AGCCAACAGCC.................................................  1374
D_ELI         ............................................................  1383
G_LBV217      ............................................................  1377
H_VI557       ............................................................  1380
O_ANT70C      ............................................................  1395
CPZGAB        ............................................................  1410
CPZANT        ............................................................  1464
A_ROD         ............................................................  1383
B_EHOA        ............................................................  1428
SD_MM251      ............................................................  1389
STM_STM       ............................................................  1392
VER_AGM3      CGAGT.......................................................  1422
GRI_AGM677    AAACA.......................................................  1398
TAN_AGM17     CCGCGCATGGATTCCCCACAGGT.....................................  1437
SAB_SAB1C     AACGGGATTACAGCAGGCCGGAGGAAAATTGGTAT.........................  1488
SYK_SYK       AGGACATCGAAGATGGGCCATGGCTCACATGGTCAGCACAGATGAGCCAACAAGCACAAGCGAAGGCACA  1541

GAG_CONSENSUS ????????????????????????????????????????????????GAga??g??g?g?a???g  1172
A_U455        ..........................................................--A-TCTTT-G-ATGGG-  1380
B_HXB2R       ..........................................................----GCTTCAG-TCTGG-  1398
C_UG268       ..........................................CCACCGGCA----GCTTCAG-TTC...  1398
D_ELI         ..................................................----GCTTC-G-TTTGG-  1401
G_LBV217      ..................................................----GCCTC-G-TTCGG-  1395
H_VI557       ..................................................----GCTTC-G-TTTGGA  1398
O_ANT70C      ..................................................---GAG-AA-T-A-GGGA  1413
CPZGAB        ..................................................----GTTAT-G-T-CCA-  1428
CPZANT        ..................................................----TCTATCA-G-G...  1479
A_ROD         ..........:.......................................--TCCA-CA-T-G-TCTA  1401
B_EHOA        ..........................................................--TCCA-CA-A-G-GAT-  1446
SD_MM251      ..........................................................--CCCA-CT-T-G-TCT-  1407
STM_STM       ..........................................................--CCCA-CT-C-G-TCT-  1410
VER_AGM3      ..................................................CCATAC--CCCT-CAAA-A-GCTC  1446
GRI_AGM677    ..........................................GCTTAC--TCCA-CAAA-A-GCTC  1422
TAN_AGM17     ..................:...CCGCCAGTTGCAGGAGCCTAC--TCCA-CCAA-AGACTC  1476
SAB_SAB1C     ..................GCAGACAGACCTCCGACCAGAGGACCAGGGCCAGAC--TCCA-CAACAGCCCT-  1542
SYK_SYK       GAACTCTCCCTCGAAGAAACCCCCCACCAACAGGGAAGTGCTCTCGCCGAAG----GCAGT-G-A-G...  1608

GAG_CONSENSUS g?ggag?a????a?????c??gg??agaagcag???gagcaga?aga?a?g??????????????????  1202
A_U455        -AAA--.....-TGACCTCCCCTGC---A---...----T--A---C-G-.....................  1422
B_HXB2R       -TA---ACA...-CAACT-CCCCTC--------...----C--T---C-A-.....................  1443
C_UG268       -A--G-.....-CAACT-CCCCTCC-------...----C--A---C-G-.....................  1440
D_ELI         -AA---.....-TAACC-CCTCTC-A--A---...-------A---C-A-.....................  1443
G_LBV217      -A----.....-TAGCC-CCTCCCC-------...----CA.....-A-.....................  1431
H_VI557       -A---A.....-TGACC-CCCCTCCA--A---...---AG--AG--TGCA.....................  1440
O_ANT70C      CA----A-T.....................---...--A----A--GGGGCCCGAAC.............  1446
CPZGAB        -A---A.............-AGA---GC---...---A---A---GGGA.....................  1461
CPZANT        -A-C-CA-GAGG-CTCAGAAG--TCTC---GG-...---G--.............................  1518
A_ROD         CT----A-ATAT-TGCAG-AA--GA-A-GA---AGA-------G---G-GACCATACAAGGAAGTGACAG  1471
B_EHOA        CTAA--A-CTAC-TGCAA-TA--GA--------AAG---A-C-G---G-GACCCTACAAG..........  1506
SD_MM251      CTAA--A-CTAC-TGCAGTTG--CA--C-----AGA--AAGC-G---G-A-CCTTACAAG..........  1467
STM_STM       CT-AGAAGTTAC-TGCAG-TG--CA--------AGA---AGC-GGA-G-CACCCTACAAG..........  1470
VER_AGM3      CT-C--C-ATATGCAGACAAG--GA--C--TT-AGG--A--A-GGA-A-AACCACCAGCAGTGAAT....  1512
GRI_AGM677    CTCC--C-GTATGCAGAGAAG--AC--CGC-T-AGA---G---G---ACA-ACAAGGAAACAG.......  1485
TAN_AGM17     CT--G-C-GTATGCAAAAAAA--AG-CC-A-TAAGGAG---A-A---G-A-.....................  1527
SAB_SAB1C     TTAA--C-GTATGCTGTT-AG--GA-ACG----.....................AAACAGCAGTGGC  1588
SYK_SYK       -A----ACC.....................................-A-.....................  1620
```

```
GAG_CONSENSUS ????????????????????????????????????????????????????gaa????????g?at?c???   1209
A_U455        ..................................................---......CA-AC-C-TTT   1436
B_HXB2R       ..................................................---......CT-T--C-TTT   1457
C_UG268       ..................................................---.....               1443
D_ELI         ..................................................---......CT-T--C-TTT   1457
G_LBV217      ..................................................--G......CTAT--C-T..   1443
H_VI557       ..................................................---.............T-T..   1446
O_ANT70C      ..................................................--G......CTCT--C-GTT   1460
CPZGAB        ..................................................--GAGCAGTCTAT--C-ACC   1481
CPZANT        ..................................................---......CTACC-C-CTC   1532
A_ROD         AGGACTTACTGCACCTCGAGCAGGGGGAGACACCATACAGGGAGCCACCAAC-......GA-G-CTTGCT   1535
B_EHOA        ..........................................GAGGTGAC-......GA-G--TTGCT   1526
SD_MM251      ..........................................GAGGTGAC-......GA-G--TTGCT   1487
STM_STM       ..........................................GAGGTGAC-......GA-G--TTGCT   1490
VER_AGM3      .....................................CCCGATTGGACA--G......GGAT--T-T..   1536
GRI_AGM677    ....................................AAGGAGAAA---GTG...GA-G--GTT..   1509
TAN_AGM17     ..................................................---CTG...GA-G--TAT..   1542
SAB_SAB1C     AAAACCAC.................TCGCCACAACAGAGCCCCTACGAG--G......GCTT-CAGC..   1632
SYK_SYK       ..................................................AGT......CTCT--C-C..   1632

GAG_CONSENSUS ????tccctcaa?TC?cTctttgg??aa?????????gaccaa                              1235
A_U455        AGTT--------A--A--------CA-CGACCCCTTGTCA--G                               1479
B_HXB2R       AACT-------GG--A--------CA-CGACCCCTCGTCA---                               1500
C_UG268       ....C--T-A-TT--C---..................A-ATC-                               1464
D_ELI         AACT--------A--A--------CA-CGACCCCTTGTCG---                               1500
G_LBV217      .......T-A-CT--C---..................A-ATC-                               1461
H_VI557       ....C-TT-A-CT--C---..................A-ATC-                               1467
O_ANT70C      TGCC--------A--C--------GAC-.........------                               1494
CPZGAB        AACC--------A--A--------CAGCGACCCCTCATC---G                               1524
CPZANT        GTAT-----G--A--C--------CA--.........------                               1566
A_ROD         G...CA------T--T--------AA--.........-----G                               1566
B_EHOA        G...CA------T--T--------AG--.........-----G                               1557
SD_MM251      G...CA------T--T--------AGG-.........-----G                               1518
STM_STM       G...CA------T--T--------AG--.........-----G                               1521
VER_AGM3      .......T-G--C--C--------AG--.........------                               1563
GRI_AGM677    ....---T-G-GC--C--------AGG-.........------                               1539
TAN_AGM17     ..............--CT-G..................AGTTCC                               1554
SAB_SAB1C     ....--------GG--T--------AG--.........------                               1662
SYK_SYK       ....--T----GC--T--------AG--.........------                               1662
```

HIV1 GAG CONSENSUS

```
          p17 ->          /<-       nls        ->/
            /<-    membrane binding     ->/
CONSENSUS-A   ATgGGTGCGAGAgcGTCagTatTaAgtGgggggaaaATTaGATgcaTgGgAgAaaATTcgGtTAaGGCCaG     70
CONSENSUS-B   -------------------------c------g--------ag------a--------------------     70
CONSENSUS-C   -------------------a-------A--c------------a-------a------A------------     70
CONSENSUS-D   -------------------------C-----------------a------A------------C-------     70
CONSENSUS-F   -------------------------C-----------------------A-----------------g-     70
CONSENSUS-G   -------------------------C---------?------?------A------?--?-G-------     66
CONSENSUS-H   ------------------------?--C--------?---------?-----------------?-------     66
CONSENSUS-O   -----------?-----T--G--G-CA--?A-?-----G-----------AC?----A------?-----     65
CONSENSUS-CPZ -----?---?-?-----?--T?-?-??--?-??--??-----???-----A-???-??--C-T?----C-     48

                                                        /<- nls ->/       p17 \/ p24
CONSENSUS-A   ggGGAAagAAAaaaTATagacTgAAaCAT?TAgTaTgGGCaaGCagggAGcTgGAaagATTcgCACTtaA     139
CONSENSUS-B   -----------------a-t-A------a----------------------A---c--------g----     140
CONSENSUS-C   -----------c-c----TGa-a-----cc-------------------------------T--------     140
CONSENSUS-D   -a----------------a---a------a----------------------a---C----T--------     140
CONSENSUS-F   --------------------?----g---C----------------------A---c----T---a----     139
CONSENSUS-G   -----------?--------A-?--?-------------------------?-?---T--------     130
CONSENSUS-H   -----------?-------?--?----------------------?--------------T-----?--     130
CONSENSUS-O   -ATCT--A--G??------?--A-----?T----------------?-------------?----TG---     129
CONSENSUS-CPZ -?--?-??----??--?-TGA-?--------?--?--------???-------?-??-?--?--??????      93

CONSENSUS-A   cCCtaGccTTtTagAaACAgcaGaagGatGTcaacaaaTAaTggaaCAgtTacAatCAgcTcTcaagACA     209
CONSENSUS-B   t---g----g---------t-------c---aG-------c---g----c-----c--t-c--tc-----     210
CONSENSUS-C   ----g----------g---t-------C---A----------aA-----C-----C-------TC-----     210
CONSENSUS-D   T---G----------g---t-------c---A----------a-g----C-----C-----a-tc-----     210
CONSENSUS-F   t---g-----c--------t-------C----g-a-------A-G----------c--T-C--TC-----     209
CONSENSUS-G   ----G----?---------A-------?------------?--??------G---C--?-----C-----     194
CONSENSUS-H   ---?G?---------?---?-------C----T---??----A------C----?C-------T------     193
CONSENSUS-O   T---G??--A?-------T-----?--TA?-G-G?-?C-GC-AC-G------G-GC--------------     192
CONSENSUS-CPZ -?-C???--??-?----?????-----?---???A-?????-?C-?--A--????C-??-??-??-A-?-     129

CONSENSUS-A   Gga?CAgAAGAAcTtAaaTCaTTAtttAAtaCagTAGcAacCCTcTatTgcgTaCAtcaaaggATAgA?g     277
CONSENSUS-B   ---t------------g--------a-----------------------T--g--------------g-     280
CONSENSUS-C   ---A----g-------g--------a---c------------t--------T------g--?a------g-     279
CONSENSUS-D   ---t---------------------?-------------------------t------g----------g-     279
CONSENSUS-F   ---T-------G----g--------a-------------gt--------tt----------g----G-     279
CONSENSUS-G   ---A----G--G?--------?---?----?--?-----??T----??--T?----------------g-     254
CONSENSUS-H   ---A-----------C--------------CT-T-----??---------?-----------A-----T?     259
CONSENSUS-O   --GT----???C--G?-?--?C-C-GG--CG--A---?-GT?----GG--?--T--CA-C--AT?T?--A     250
CONSENSUS-CPZ ?-?T-????-?-?-????--?--G-----C--??-????GT?--G-GG---?-----A??G?????A?-?     168

          /<-     MHR      ->/
CONSENSUS-A   TAAaaGACaCCAAgGAaGCttTagAtAaAaTaGAgGaAatacAAAatAAgagcaagcaaaag????????     339
CONSENSUS-B   ------------------------?--g---------GAg-----C--a--t---a----?........     340
CONSENSUS-C   --Cg---------------c-----C--g---------GA------C--a--TC-------A------..     341
CONSENSUS-D   ------------------------a--g--g------GA------C--A--T---A-------------     341
CONSENSUS-F   ----G---------------------c--gC--------GA------C--A--Tc--------.......     341
CONSENSUS-G   -------------A------c----GG--G-G--aA-??c------G--c--TC-------A........     314
CONSENSUS-H   ----------------------???--G?--?----?G?------C--?--?C--?-?-??........     308
CONSENSUS-O   -??G---T--?C--C-G--AA--C-A--GT--A-----G--...-TGG??----G-A-GTCT........     304
CONSENSUS-CPZ --?-------A?-??----??-??-A????-?A-???G?A????????-???AC-?????-AAC-G-AAG     203

                                  p24 \/        \/   'p2'   \
CONSENSUS-A   ??????????acacagcaggCaGcaGCT......g?cACA...ggaAgC???......agcagcaa?gtc     379
CONSENSUS-B   ..........g-------a---------??????-a-----...----a----??????------c-g---     382
CONSENSUS-C   ..........------------aa---g...gct-A-???...----Ag...................---     374
CONSENSUS-D   -.........g-------A---a-----......-A-----...a---a-.........------C-G---     383
CONSENSUS-F   ..........--------a-?g------......-A--A-...--g.....................---     370
CONSENSUS-G   ..........-?----------??????......-A??A-...----A-.........------C-A---     347
CONSENSUS-H   .........--------?---A---?-.......-AT-A?...-?--A?.........G?--A?--G?--     342
CONSENSUS-O   .GC-GA-GCCG-TA--G-A..............-A?---......---.........?C?--GC-G?C?     336
CONSENSUS-CPZ TA-C--A---????G???A-G??????????C???-?????????????---??????????????-???     212

                                  Zn-motif                              pol cds ->
                        / p7          /<-Zn-motif ->/      /<-Zn-motif ->/ p7 \/  'p1'
CONSENSUS-A   ???????????AGcCAaAAtTACCCtATAGTgCAaAAtgcacAagGgCAAatgatAcatCAG?cctTaT     436
CONSENSUS-B   ...........-----------------------g--ccTc--G---------G--------g--a---     440
CONSENSUS-C   ...........--T--------t-----------g---CTc----------g----c---g--a---     432
CONSENSUS-D   ...........-----------t--------------CcT---g---------G----c---G--a---     441
CONSENSUS-F   ...........--T-----------------A--g---cTt--G--a------G----------a---     427
CONSENSUS-G   ...........--?---------t----------G-----------------?----c------a---     402
CONSENSUS-H   ...........--T--------T--?-----A--?-----C-----?--G??-G----C---G--A---     395
CONSENSUS-O   ...........?-T-----------------A?C------?--G--A------G-------?G--A-C-     390
CONSENSUS-CPZ ------------?-??G---------C?-C?-A????----???---?????????G--------??-??     244
```

```
                                  vpr binding                (minor) \/ vpr binding     p6 terminus
          'p1'          \/ p6  /<-->/     \/ (minor)                    /<- ->/     / (80%)
CONSENSUS-A   CACCtAggACttTgaAtGCATgGGTgAAaGtaaTAGAagAaAagGcTTTcAgcCcAGAaGTaATACCCAt    506
CONSENSUS-B   -------A-----A----------A------g-------g-------------------------------    510
CONSENSUS-C   -------A----------------a----------g--g----------------g-----------    502
CONSENSUS-D   -------a--------c--------a---------------?-------------------------------    510
CONSENSUS-F   -------A-----A----------a--g--g--------g--------t--------------------    497
CONSENSUS-G   -------?----------------A------g-------------C-----T-----------------    471
CONSENSUS-H   ----------C--A----------A------G-------?--------T-----------------?--    463
CONSENSUS-O   -C--C---------A----------A--G-C-G-------?--?--C--T-A---T---A-?--T--T--    457
CONSENSUS-CPZ ----?--A--C--?-----?-----A--????G----?--?--????--?-??--?--?--?--?--?--    293

CONSENSUS-A   GTTctCaGcatTaTCAGAgGGaGCcACccCaCAagAttTaAAtatgATGcTaAAcatAGTggggGGaCAc    576
CONSENSUS-B   ---T--------------A-----------------------c-Cc----------C------------t    580
CONSENSUS-C   ---ta-------------a-----------------------c-cc---t----t-c------------t    572
CONSENSUS-D   ---T--------------A-----------------------c-CC----------C------------T    580
CONSENSUS-F   ---t--------------a--g-----t--------------c-CC--------t-C------------T    567
CONSENSUS-G   ---T--------------a-----------------------cC--------?-C---------g--T    540
CONSENSUS-H   ---T--------------A-----------------?------GCC---?----T-?---------?--T    529
CONSENSUS-O   ---?ATG-----?-----?--?--??T??-CT-T---A-?----C?-----?--TGCCA-A--?-----?    514
CONSENSUS-CPZ ---T--?--T?-------?--?--????--T--?---?-?--C-CC---?-?--TGC---?----??--T    347

CONSENSUS-A   CAgGCAGCtATGCAaATGtTaAAagAtaCcATcAATGagGAaGCTGCAGaaTgGGAcaGg?tACAtCCAG    645
CONSENSUS-B   --a-----c-----------------g-------------------------------T--At-g-------    650
CONSENSUS-C   --a-----c-----------------------------g-------------t---t---------    642
CONSENSUS-D   --a-----c-----------------g-------------------------t---c---------    650
CONSENSUS-F   --a-----c----------------c-------------------------------aT---------    637
CONSENSUS-G   --A----------------c----G-----t--------------------G---------A-------C    610
CONSENSUS-H   --A-----C--------??----------A-------------------------T-------------    596
CONSENSUS-O   --A-G?---T-A---G--?-?--G--AGTA--------------A-?---?-----T--AACT--?---C    578
CONSENSUS-CPZ --A-G---C-----?G--C----?--AGT?--?-----------?--?--G-----T?-?T----?--CA    408

CONSENSUS-A   TaCatGCAgGGCCTattcCaCcagGcCAgaTgAGaGAACCAAggGGaAGtGAcATAGCAGGaacTACTAG    715
CONSENSUS-B   -g---------------g-------------------------------------------------    720
CONSENSUS-C   -?------------?--g----------a-------------------------------------    710
CONSENSUS-D   -g---------------G-------------------------------T-----------------    720
CONSENSUS-F   -G--g-----------C--------------a--g-----T----------t-----t----------    707
CONSENSUS-G   ?---g---------------------------A-----?--------------T-------------    678
CONSENSUS-H   -G-------------------------?-------------------?--T-----------------    664
CONSENSUS-O   C-?C??T?-----GT-?--------G-----A--G-------CA----------T--T--?--A-----    642
CONSENSUS-CPZ CT-----?--?--??-A????----A--??-A--G--?----??--?-----T--------G--?--?--    462

CONSENSUS-A   tAcCctTcaaGAAcAAaTaggaTgGATgACaagc....:.....AAtCCACCtatCCCaGTgGGagAcATc    776
CONSENSUS-B   ---------G-----------------aT.........----------------a-----a---    781
CONSENSUS-C   ---------g----------c----------T.........--C--------T---------------    771
CONSENSUS-D   ---------G----------?-------------.........-----------------A-----A---    780
CONSENSUS-F   ---------G---------ca-------------.........--c------g-------------a---    768
CONSENSUS-G   ------?--G---------A----------C---.........-----------------------a--t    738
CONSENSUS-H   ---------G----------C----------?--.........-----??-------------------    722
CONSENSUS-O   C--?-AG-----G-----T???----CT--?--?...????C?--C????----------A---------    694
CONSENSUS-CPZ C--A?-G--G--?--?--????----???--?CA??????????---???????-------A--G--??-?    502

CONSENSUS-A   TATAaaAgaTggATAaTccTgGGatTAAATAAaATAGTaAGaATGTATAGcCCTgttAGCATttTgGAca    846
CONSENSUS-B   ----------------------------------------------------acc------c------    851
CONSENSUS-C   ----------------T-----G-------------------------------C-------------    841
CONSENSUS-D   ----------------------------------------------------c-------------    850
CONSENSUS-F   --------------c-----a-----------------------------------C-------------    838
CONSENSUS-G   ----------------------------------------------------c-------------    808
CONSENSUS-H   ---------------------------?-----------------?-----C-------------    790
CONSENSUS-O   ----G--A-------G-G?-A---C----C-----G----A------C--?--A--G-----C--A--T-    762
CONSENSUS-CPZ ----??------?-?--??-?--?-----?--?G-G--?--?-?----?-?--A--?--?--?C-?--?-    552

CONSENSUS-A   TAAgaCAAgggCCaAAaGAaCCcTTcAgaGAcTATGTagAtAggTTCTtTAAaactcTcAGAGCtgAgcA    916
CONSENSUS-B   ----------a-----g--------T--------------cc------A---------A-----c-----    921
CONSENSUS-C   ---a------------G--------T--------------C-------------T-A--------A--    911
CONSENSUS-D   ----------A-----g--------T--------------?C------A---------A-----c-----    919
CONSENSUS-F   ----------------------T--------------C----------c--A----------    908
CONSENSUS-G   ----------------??????????????????????????????????????????????????    824
CONSENSUS-H   ---?------?--------------T--------------?------------?--T-A--------?--    855
CONSENSUS-O   -T-?G--G--A-----?-----A--?-----?---------C----?-AC-----AT-A----------    827
CONSENSUS-CPZ -?-?------?-----?--G-----?-----T-----G--?--A----A?-----??-?--?--?--?--    609

CONSENSUS-A   aGCTaCaCAgGagGTAAAaaacTggATGACaGAaACctTgcTggTCCAAAAtGCgAAtCCAGAcTGTAAg    986
CONSENSUS-B   ----t----------------t-----------------t----------------c-----t------    991
CONSENSUS-C   ---------A--T--------t-----------c------T----------------C-----T------    981
CONSENSUS-D   ----t-------t--------t--------?---------T----------------C-----T-----a    988
CONSENSUS-F   ------------A-----gggT----------C------T----------------C-----T------    978
CONSENSUS-G   ????????????????????????????????????????????????????????????????????    824
CONSENSUS-H   ---?---------?--?--???T----------C------T-----------------------T--?---    918
CONSENSUS-O   ------T--A--A-----?--T--------------??----T--T--?---?-?--?-----T--?--A    889
CONSENSUS-CPZ ?---T-----???--G---???--------?--?--??-?T-A?-?--?-----?--?-----T--?--?    660
```

```
CONSENSUS-A   tCCATttTaAgAgcatTagGa?cagggGCtaCaTTAgAaGAaATGATGACAgCaTgcCAgggAGTggGaG  1055
CONSENSUS-B   A-t-------A------g---C---ca------c----------------------T-------------  1061
CONSENSUS-C   A--------------------C--------t-------------------------T-------------  1051
CONSENSUS-D   A-t--C----A------g---C---?----------------------------t-----------g-  1057
CONSENSUS-F   A---------A------G---C----------c----------------------t-------------  1048
CONSENSUS-G   ????????????????????-????????????????????????????????????????????  824
CONSENSUS-H   A?T-------?----------?-------T--A--------------------?--?-----------  982
CONSENSUS-O   CA?---?-G-A-?--------?C---?A-----?--------?------GT---C--T--?-----A----  951
CONSENSUS-CPZ CA?--?C-?-A?--?--G-----?--?--C?-C?----?------?-?-----T--T--A-----A----  717

CONSENSUS-A   gACCcgGcCAtAAaGCAagggTtTTgGCTGAgGCAATgAGtCAaGta.........caacAt????caAa  1112
CONSENSUS-B   -------------------A-----------a--------C------.........ac-a-----t--..  1117
CONSENSUS-C   ----Ta----C--------A--G----------------C----C-.........A-CA--...a----  1109
CONSENSUS-D   -----a-t-----------A--------------------C----c-.........ac-a---...t--?-  1114
CONSENSUS-F   ----t--------g-----a--------------------c----c-.........ac-a--...-????  1101
CONSENSUS-G   ??????????????????????????????????????????????.........??????...-????  824
CONSENSUS-H   -?--T?----?-------?A------------------C------???......AC-A--...-????  1031
CONSENSUS-O   -?--AACT--?--G--?-?A?-?C-A--A--A------GC-?C?-CCCAGCAAGATTT?A-AGGAGG-T-  1012
CONSENSUS-CPZ -?--A?C?-----G-----?--??-?--?--A--?---?C????-?T?A????...???--?...-???G  753

CONSENSUS-A   c??????...aTAATGATgCAgAgaGGcaAttTt...agggGccagaaaaga???aTtAAgTGtTTcAAc  1167
CONSENSUS-B   .gctacc???----------...---aa---a-g--Agactg----------t                    1180
CONSENSUS-C   -.........--------------a--------...-Aa----Ct------AtTg----A---------  1167
CONSENSUS-D   T-ctgC-...-------------------...-A-----cA-g--a-att---------------------  1176
CONSENSUS-F   ?gca-cc...-----------a-a-t--c---...-A------A-G----AtTg----A-----t--T  1163
CONSENSUS-G   ?------...--?--------?-A-?--???---...-A-----C--?--?-A-T----------?--  871
CONSENSUS-H   ?GCAG--...--?----------A---?--C--?...-A---?--A-G----ATT?----A--------?  1086
CONSENSUS-O   -ACAGCA...G--T-C-----A-----GC-?AA-...CCAA?TAG?---G-?-C---A--A--?-----?  1068
CONSENSUS-CPZ A-C-G--...G-?T-??-?----??--??-?GG???????--???A---------?-?--A--?--T--?  790

CONSENSUS-A   TGTGGcAaaGAaGGaCAccTAGCcAgAAATTGcAggGCcCCTAggAAAAagGGcTGtTgGAAaTGtgggA  1237
CONSENSUS-B   ----------g---a------a-------------------------------------------A-  1250
CONSENSUS-C   --------g-----G---a-------------------------------------------------A-  1237
CONSENSUS-D   ----------g---a----a-a-----------------?----------------------A-  1245
CONSENSUS-F   ----------------a------a----------------A---------------------a-  1233
CONSENSUS-G   --------G--------T-----------------------?-----------------A-  940
CONSENSUS-H   ----------?----?-----------------------?-----------?--A-  1152
CONSENSUS-O   -----A-----G-----TA----A-?---?--TC-A--A-----?----?---T--C----------AC  1134
CONSENSUS-CPZ --?--?-----?--?--????--?--------?-A?--?--A--A-G?--A-----?---?G?-----?C  843

CONSENSUS-A   agGAaGGaCAcCAAaTgAAagAcTGcACt...???gAg...aGaCAGGCtAATTTTtTAggGAaAaTtTg  1298
CONSENSUS-B   ----------t--t---...------???----------g--c--  1311
CONSENSUS-C   ?----------T---.....---...--g----------  1297
CONSENSUS-D   ----------t------.....--a...----------  1306
CONSENSUS-F   ga----------.....--a...----------  1294
CONSENSUS-G   ?---?-----T----------A.....---...----------?---?--  997
CONSENSUS-H   ?----------T--G?-------?-----A.....--?...----------?--  1208
CONSENSUS-O   -------T----------T----?A.....A-TGGA-?------A----------?--?TAC--  1194
CONSENSUS-CPZ -?--?--------??-?---?----??-A???---???????--?-?-T?----------?????  884

CONSENSUS-A   GcCTTCCagCAAgGGg...AGGCCagGaAAtTTtCctCAgagCAGacca.....  1344
CONSENSUS-B   -------ca------a...--------g-------t----------??????????????????  1357
CONSENSUS-C   -------cA-------...--------G-----C-T---------g---??????????????????  1343
CONSENSUS-D   -------ca------a...--------g--c--?-t----------.....  1351
CONSENSUS-F   --------A-------...-----?--?-------T?----------.....  1337
CONSENSUS-G   -------CA-------...-----?--G-------T-----?---??--.....  1039
CONSENSUS-H   -?--?---?---A---...--------G-------T?--------G---.....  1250
CONSENSUS-O   ----C-GG-GGGCAC-...--------C----A-GTG----?AC?-G??.....  1236
CONSENSUS-CPZ ?--??-??-??G?---???--?-----?--C---GTG----A??????-??????.....  914

CONSENSUS-A   .....GAgCCaaCAGCcCCACcaGCAGAg.....a?ct  1371
CONSENSUS-B   ??????.....????????----------a----.....-g--  1385
CONSENSUS-C   ??????????????????----------??????????????????-G--  1371
CONSENSUS-D   .....----------.....-G--  1379
CONSENSUS-F   .....----------.....-g--  1365
CONSENSUS-G   .....------?----------?-----?.....?-??  1060
CONSENSUS-H   .....----------.....-G--  1278
CONSENSUS-O   .....??C---T----------?ATG---.....  1257
CONSENSUS-CPZ .....--?-----?--?-----?AT----.....--?-  936

CONSENSUS-A   tt???gGgatgggggaagagataacc...tCctct.....ccgaagcaGgAgCagaaagac..  1421
CONSENSUS-B   -C...a--T-t----------c---aactc-----???????????-a----------c--t----..  1438
CONSENSUS-C   -C...a--T-C...--G----C----...C--g--.....----------c-------??  1418
CONSENSUS-D   -C...---T-t-----g---------...C-----.....-a---a----------??  1429
CONSENSUS-F   -c...---T-C?-a--G---------...C-----.....----------?---..  1413
CONSENSUS-G   ??...?????????--G------?--...?----?.....----------?-C???????..  1088
CONSENSUS-H   -?...---T-?--A--G--?--G---...C--?--.....--?--?------???-?G--?..  1317
CONSENSUS-O   .....--G-?AG-G-AG...G?ACA?.....GA---T---???-----?-GG..  1289
CONSENSUS-CPZ A-...??????CA---G--??A?-AG...AG????.....?A----?????????G?G???..  959
```

```
CONSENSUS-A   ....??????aaggaacag??tcc?cCtTtAgtttCCcTCAAATCACTCTTTGGCAACGACCccTtGTCA   1478
CONSENSUS-B   .......----------t-TA-??--------C--------g----------------------c----   1495
CONSENSUS-C   ????-------g----???--???---C---Ac-------------------------------       1464
CONSENSUS-D   ????-------------t-ta-...-------C-------------------------------------   1486
CONSENSUS-F   ..........g---g--t-ta---T--c----C-------------                          1449
CONSENSUS-G   ..........-----?-TATA-??-------??-------------                          1118
CONSENSUS-H   ..........???---.....?-T--?---?C-------------                          1341
CONSENSUS-O   ....G--C-G?-?--G-T?TA?...--?--T-CC----------C--------G-CA----AA-A-     1341
CONSENSUS-CPZ ..........??????-TA--?-----???????-----?-----?----------??----??-?A--C  995

CONSENSUS-A   CAG                                                                    1481
CONSENSUS-B   --aTAA                                                                 1501
CONSENSUS-D   --A                                                                    1489
CONSENSUS-CPZ ---                                                                    998
```

HMMER Sequences in the Pol Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_HXB2R	HIVHXB2R	K03455	Starcich,B.	Science 227, 484 (1985)
D_ELI	HIVELI	KO3454	Alizon,M.	Cell 46, 63 (1986)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68, 1586 (1994)
O_MVP5180	HIVMVP5180	L20571	Gurtler,L.G.	JVI 68, 1581 (1994)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
CPZANT	SIVCPZANT	U42720	Vanden Haesevelde,M.	Virology 221, 346 (1996)
A_ROD	HIV2ROD	M15390	Clavel,F.	Nature 324, 691 (1986)
B_EHOA	HIV2EHOA	U27200	Rey-Cuille,M.A.	Virology 202, 471 (1994)
SD_MM251	SIVMM251	M19499	Franchini,G.	Nature 328, 539 (1994)
STM_STM	SIVSTM	M83293	Novembre,F.J.	Virology 186, 783 (1992)
VER_AGM3	SIVAGM3	M30931	Baier,M.	Virology 176, 216 (1990)
GRI_AGM677	SIVAGM677	M66437	Fomsgaard,A.	Virology 182, 397 (1991)
TAN_AGMT17	SIVAGMT17	L19254	Hirsch,V.M.	Virology 197, 426 (1993)
SAB_SAB1C	SIVSAB1C	U04005	Jin,M.J.	EMBO J. 13, 2935 (1994)
SYK_SYK	SIVSYK	L06042	Hirsch,V.M.	JVI 67, 1517 (1993)

The following alignment was generated using the HMMER program as described in the introduction to this Part and in Part III. For simplicity, only representative types and subtypes are shown. An ordinary consensus sequence (lowercase signifies majority, uppercase signifies 50% or greater) was constructed from these representatives using MASE; this is not a "most likely sequence" derived from the HMMER model (Part II). Annotation is based on HIV-1s, therefore the user should be cautious about its applicability to other PIV sequences.

```
POL_CONSENSUS A??????AGaag???gagatgttggaAatgtGgacA????g?a??????????????????????????????     29
B_EHOA              ---GCA-G----C--------------GCAA-G-CA.........................     36
SD_MM251            GT-TTG--AT---G----GGAG-----CTCT-C-AGGCAATGCAGAGCCCCAAGAAGACAG     61
STM_STM       -GGCCCG-----ACAAG-T-----------------ACAG-GCCA.........................     45
VER_AGM3      -CC...A----AGAT------C-T---------GA-...ACC-GG.........................     39
GRI_AGM677          GT-CGGC-A-AT-G-CC-TATG-C-A-.................................     27

POL_CONSENSUS ????????????????????????????????????atggcaaAatgcccagaa?GACAGGc?ggTTTTTTAg     64
A_U455                                                                  --------      8
B_HXB2R                                                                 --------      8
D_ELI                                                                   --------      8
O_ANT70C                                                                --------      8
O_MVP5180                                                               --------      8
CPZGAB                                                                  --------      8
CPZANT                                                                  --------      8
A_ROD                                                          ------A----------     17
B_EHOA        ..............................CATC---T--------------A-------G----------     77
SD_MM251      GGATGCTGGAAATGTGGAAAAATGGACCATGTT-----C-----------CA-------G----------    131
STM_STM       ............................TCAG-----C------------A------TG----------     86
VER_AGM3      ............................GCATT-A------GATTGCAG-G------TAAA--------     80
GRI_AGM677    ........................................-G-C---AAGA-TG-------AAA--------     59
TAN_AGMT17                                                  --G-G-------CAA-------A     23
SAB_SAB1C                                                               --------      8
SYK_SYK                                                                 --------      8

POL_CONSENSUS Ggaa????ggccgg???????????gggaag???ga?gcc?agcaattt????????????????????     93
A_U455        ------...AATTT--CCTTCCAACAA---G--...-CCAGGG-ATTT.....................     48
B_HXB2R       ------...GAT-T--CCTTCCTACAA------...-CCAGGG-ATTT.....................     48
D_ELI         ---G...AATTT--CCTTCCCACAA------...-CC-GGG-A-TT.......................     48
O_ANT70C      -C--...ATA-T--CCTCC...GGG---C-C...--G---AG------A.....................     48
O_MVP5180     ------...GTA-T--CCTCC...GGG---C-C...--G---AG------A.....................     48
CPZGAB        ------...A--TT--CCTTCCCGCAGC--G--...-CCAGGC-A-TT.......................     48
CPZANT        ----ACCGAC--CCACGTG...GTG---GT-...C-GA--AG-G--C--TGTGCA................     57
A_ROD         -ACT...G----CTTG.........-----A...--A---CC----C--CCCCGT................     57
B_EHOA        --TT...C----CCTG.........-----A...--A---TC-------CCCCGTCCAGGCACCCCAGGG    132
SD_MM251      -CCT...T--T-CATG.........-----A...--A---CC-------CCCCAT................    171
STM_STM       -CTT...T----CATG.........-----A...--A---CCA---C--CCCCAT................    126
VER_AGM3      --T-...T------TGGAT.........-----GC...--AA--C--A-----TCCCGC.............    123
GRI_AGM677    --T-...T----ATTG.........-----G-AGC--AA--A--A-----TGTGCAATACAG.........    108
TAN_AGMT17    ---G...AATG-CCGCACCCTAAGGTTCC-A...-CC--GC-ATTT---AGAGCAAGGAGG.........     78
SAB_SAB1C     --TT...T----CCTGGGG......CA--G-...--AA--C--G-----.....................     45
SYK_SYK       ------...A--AT--GGGGT...CTCCA-G--...-CCC--AGC---C--CCCAGTGAGGAGCGAGCCATC     69

POL_CONSENSUS ????????????????????????????????????????????????????????????????????t     94
A_U455        ..........................................................................-     49
B_HXB2R       ..........................................................................-     49
D_ELI         ..........................................................................-     49
O_ANT70C      ..........................................................................-     49
O_MVP5180     ..........................................................................-     49
CPZGAB        ..........................................................................-     49
CPZANT        ..............................................................GAAGGAGGAA     67
A_ROD         ..............................................................GGCCCAAGT-     67
B_EHOA        GATAGTGCCATC.....................TGCGCCCCCGATGAACCCAGCATTCGGCATGACA    178
SD_MM251      ..............................................................GGCTCAAGTG    181
STM_STM       ..............................................................GGCCCAAATA    136
VER_AGM3      ..............................................................CGCTACTCT-    133
GRI_AGM677    ............................................................AGGAGACACAGT-    121
TAN_AGMT17    ..............................................................AGCCGTTCCGACG     91
SAB_SAB1C     ..........................................................................C     46
SYK_SYK       TGCTCCCCCTCTGGAGGACATCGAAGATGGGCCATGGCTCACATGGTCAGCACAGATGAGCCAACAAGCA    139
```

HIV1 POL

```
POL_CONSENSUS c?tcagag?????????gccaacagcccc?ccag?agag???g?tg??????????gga???g???aa??   130
A_U455        -C------CAGACCAGA-----------A----C---AATCTT--GGAT...GGG---AAA-ATG-C..   114
B_HXB2R       -T------CAGACCAGA-----------A----A----AGCTTCAGGTC...TGG--TAGA-ACA-CAA   116
D_ELI         -TC--A--CAGACCAGA-----------A----C----AGCTTC-GGTT...TGG---AGA-ATA-C..   114
O_ANT70C      GTG-----ACCAGCACAC---T-------A--GATG---GAG-AA-T.....                    96
O_MVP5180     GTG----AACAAGTGTCC---T-------A---ATG---GAG-CA-T.........-A-GGAACAAG-GA   110
CPZGAB        GTG----ACAGAACAGAA-----G-----A--GAT----AGTTA--GGTACCAGGA---AGA.........   111
CPZANT        GTAGT-GA.........---------T--A--CAT----ATCTA-CA.........---GGA-CAC--..   117
A_ROD         -CG---G-...GCTGACA--------A--C----TG--TCCA-CA-T.........---TCTACTGG-..   123
B_EHOA        -C----G-...TGCGATT---T-T--A--C--T-C---TCCA-CA-A.........---GAT-CTA--..   234
SD_MM251      -A----G-...GCTGAC------T--T--C----AG--CCCA-C--T.........---TCT-CTA--..   237
STM_STM       -C----G-...GCTGAC------T-----T----A---CCCA-C--C.........---TCT-CTG-G..   192
VER_AGM3      GGGGT-GA.........------T--G--C--TCC-CC-AGTCCATA.........C--CCCTGCA--..   183
GRI_AGM677    GG--T-GA.........A----------C---ATG--AACA-C-TA.........C--TCCAGCA--..   171
TAN_AGMT17    GC--CCCCGATCGCCGC--ATGG-TT---CA---GTCC-CCA-T--C.........A-GAGCCTACG-TC   152
SAB_SAB1C     -C--T--CCTCCATCAGA--------T--C---ATG--ACGG-A-TA.....                    93
SYK_SYK       -AAGC--A.....--CACA-AACTC..                                              159

POL_CONSENSUS ????c???a?gcagcagga?aa??????????????????????????????????????????gca   146
A_U455        .CTC-CCTGC-A-A-----.....--T                                              135
B_HXB2R       CTCC-CCTCA-A-------.....--C                                              138
D_ELI         .CCC-TCTCAAA-A-----.....---                                              135
O_ANT70C      .......G-A-GGA-----G--TCA.....GGAA--                                     120
O_MVP5180     ATCAGAGTCA-A--GG---.....T--                                              132
CPZGAB        .......G-A-AGC-----G--.....GAAAG-                                        132
CPZANT        .GAGGACTCA-A--GGTCT.....CA-                                              138
A_ROD         .GAAATAT-T------A-GG--.....AAGA--                                        150
B_EHOA        .GAA-TAC-T---A-TA-GG--.....GAA---                                        261
SD_MM251      .GAA-TAC-T----TT--GC--.....GCA---                                        264
STM_STM       .AAGTTAC-T-----T--GC--.....GAA---                                        219
VER_AGM3      .GAAGCTCCT------AT-.....                                                 201
GRI_AGM677    .GAAGCTCCTC------T-TGCAGAGAAGGGACAGCGCCTGAGAGAGGAGAGAGAACAGACAAGGAAA--  240
TAN_AGMT17    CAGC-AAG-GA-TC-T--GGC-GTATGCAAAAAAAGG.....AGACCAACTAAGGAG---            207
SAB_SAB1C     ....-AGC-G--C-G----A--.....TTG-T-                                        117
SYK_SYK       .TCC-TCG-A-A-A-CCCCC-CCAACAGGGA.....AGTGCTCTC--C                         201

POL_CONSENSUS gaaaGa??????cag???g?aac?????????????????????????????????????????????   159
A_U455        ------.....---...-G---A.....                                              150
B_HXB2R       --T---.....--A...-G---T.....                                             153
D_ELI         ------.....--A...-G---T.....                                             150
O_ANT70C      -----GGGGCCCG-A...CG-G-T.....                                            141
O_MVP5180     -G----GCT.....                                                           141
CPZGAB        -GG---.....G--...CAGT-T.....                                             147
CPZANT        -GGG--.....GGA...-G---T.....                                             153
A_ROD         --G---GCA...G--AGA-AG--CATACAAGGAAGTGACAGAGGACTTACTGCACCTCGA.....        207
B_EHOA        ---G--GAA...---AGA-AG--CCTACAAGGA.....                                   291
SD_MM251      --G---AAG...---AGA-A-G-CTTACAAGGA.....                                   294
STM_STM       --G---GAG...---GAA-AC--CCTACAAGGA.....                                   249
VER_AGM3      TGC---.....--AGGG-A-G-AGTTGAGGGAACAAAGGAAAAAACCACC.....                  246
GRI_AGM677    ---G--.....G-A...AG--GT.....                                             255
TAN_AGMT17    A-----.....G-A...-G---T.....                                             222
SAB_SAB1C     TGC---.....---...ACCT-CGACCAGAGGACCAGGGCCAGACGATCCAGCAACAGCCCTGTTAAAG   178
SYK_SYK       ---G--.....G--...CAGTGGGAAGGAGGA.....                                   225

POL_CONSENSUS ????????????????????????????????????????????????????????????????   159
A_U455        .....                                                                    150
B_HXB2R       .....                                                                    153
D_ELI         .....                                                                    150
O_ANT70C      .....                                                                    141
O_MVP5180     .....                                                                    141
CPZGAB        .....                                                                    147
CPZANT        .....                                                                    153
A_ROD         .....GCAGGGGGAGACACCATACAGGGAGCCACCAA                                    239
B_EHOA        .....GGTGA                                                               296
SD_MM251      .....GGTGA                                                               299
STM_STM       .....GGTGA                                                               254
VER_AGM3      .....AGCAGTGAATCCCGATTGGA                                                266
GRI_AGM677    .....                                                                    255
TAN_AGMT17    .....                                                                    222
SAB_SAB1C     CAGTATGCTGTTCAGGGGAAACGGCAGAAACAGCAGTGGCAAAACCACTCGCCACAACAGAGCCCCTACG   248
SYK_SYK       .....GACCA                                                               230
```

```
                                  \/ prot                                                    <- gag cds end
POL_CONSENSUS ?ggag?cttt??c?tccctcaa?tc?ctCTTTGGaaaaGACCa?T?gtcacagtaaa?aTagagGG?ca?    217
A_U455        .-ACTC----AGTT--------A--A--------C--C----CC-T----------A----GA--A--G    219
B_HXB2R       .-T-TC----AA-T-------GG--A--------C--C----CC-C------A----G----G---G--A    222
D_ELI         .-T-TC----AA-T--------A--A--------C--C----CC-T---G--A----A----G---A--G    219
O_ANT70C      .CT-TC-G--TG-C--------A--C--------G-C------A-A--T----C--GAG-T-G---C--C    210
O_MVP5180     .-T-CC-A--TG-C--------A--C--------G-C------A-A-------C---GG-T-G---T--T    210
CPZGAB        .AT-TC-ACCAA-C--------A--A--------C-GC----CC-CA--C-------AG----A--G--A    216
CPZANT        .ACCTC-C-CGTAT-----G--A--C--------C--------A-GA-GGA---TCTC---C----A--A    222
A_ROD         CA---GAC--GCTGCA------T--T----------------G-A-------C-T-C--T-----T--G    309
B_EHOA        CA---GA---GCTGCA------T--T---------G-------G-A----A--C--CT--T-----T--A    366
SD_MM251      CA---GA---GCTGCA------T--T---------GG------G-A-----T-CTC-T--T--A--A--G    369
STM_STM       CA---GA---GCTGCA------T--T---------G-------G-A-------CCC-T--T-----T--G    324
VER_AGM3      CA---GGA-ATT-T-TGAA-TCCCT...-------G-------A-AAAA-----TT-C-----A--GGTC    333
GRI_AGM677    .----GA-G-TT-C-TGAG-TCCCT...-------GG------A-GAAACG---C-TC-----A--AACG    321
TAN_AGMT17    .----GA--ATT-C-TGAGTTC                                                    243
SAB_SAB1C     A----G---ACAGC-------GG--T---------G-------A-GAAA--T--GT-C--------A--G    318
SYK_SYK       A-AGTCTC-ATC-C--T----GC--T---------G-------A-GA-AGA----G-TG----A--TG-C    300

POL_CONSENSUS c?a?t?gaagcttTatTaGAtACaGGaGCaGAtGAtaCagTa??????gtaga?g??ata?a???????t    267
A_U455        -TGA-A------C-----------------------------CTTA...-A---CATA-ATTTG......C    280
B_HXB2R       -T-AAG------C-----------------------------TTA...-A---AATG-GTTTG......C    283
D_ELI         -T-AAG------C-----------------------------TTA...-A---AATG-ATTTG......C    280
O_ANT70C      -T-TGT----T---GC-G----------------C-------.....C--A-CAAC---C-A.....-    268
O_MVP5180     -T-TGT--G------C-G--------G---------------.....T--A-TAAC---C-A.....-    268
CPZGAB        -T-TGT--------GC-------------T------------.....A----GAGA---C-A.....-    274
CPZANT        AA-TGTC-G---C----------------T-----C------.....-----G-GA--TC-T.....-    280
A_ROD         -C-G-A----TC--G-----C-----G--T--C--CT--A--.....----CA-GA---G-G.....-    367
B_EHOA        TC-G-A----TA---C----C--------T-----CT--A--.....----CA-GG---G-A.....-    424
SD_MM251      -CTG-A----TA-----G--------G--T------T-TA-T.....---ACA-GA---G-G.....-    427
STM_STM       -CTG-A----TAC----G--C-----G--T--C--CT--A--.....--G-CG-GG---G-G.....C    382
VER_AGM3      -CCA-CAG---A-----------G--G----------CA-TATAAAA-A--CA-ATT--C-A.....-    397
GRI_AGM677    -C-G-GC----C--G--------------------C--TA--ATTCAA-A-A-G-ACT-GC-CTTTCCCC    391
SAB_SAB1C     AA-G-CACT---C-------C-------------CT-----.....A--C-G-GT---G-A.....-    376
SYK_SYK       -TGG-GC--ATG---G----------G----------TA--ATTAGA-A---A-AT--TC-A.....-    364

POL_CONSENSUS ta???ggaaaa???TggaaacCaAaaataaTaGG?GGaaTaGGaGG?tttaT?aa?gtaaaagaaTataa    327
A_U455        C-...------...--------------------G-----T-----T-----C--A----G-C-G---G-    344
B_HXB2R       C-...----G-...--------------G-----G-----T-----T-----C--A----G-C-G---G-    347
D_ELI         C-...-------...--------------G-----G-----T-----T-----C--A----G-C-G---G-    344
O_ANT70C      -GGAA------...--------------G-----G--T--------T-----A--A----------G-    335
O_MVP5180     --GAA----G-...----C----------G-----G--T--------C-----A--A--------G-----    335
CPZGAB        --CAA---CTT...---------------G-----A-----T-----T-----C--A--C---C---T-G-    341
CPZANT        -GCAA-----C...-----G--------C---T--T-----T--G--T-----TTCC----C--C-G-----    347
A_ROD         --...--G--CAAT-AT-GC---------G----G--------G--A--C--A--TACC---G--------    434
B_EHOA        --...--C-GCAAT-AC-CC---------G----T--G--------A-----A--TACC--T-----C--    491
SD_MM251      --...--TCC-CAT-AT-CC---------G----A----------T-----T--TACT--------C--    494
STM_STM       --...--GCT-CAA-AC-CC-----GG--G----T----------A-----A--TACT--G----T---    449
VER_AGM3      --TCA----C-...--------------------G--C--T-----GGGAC-C--T--------G----G    464
GRI_AGM677    C-...CAT---CCA---CGTT-C--GG--G----A--T--------AGGG--TC-T--C---------C-    458
SAB_SAB1C     --...---G-CAAT----------G------T--A----------A-G---T--T-------C---CC-    443
SYK_SYK       --...CACC--CCA---TC----------G----A---T-------AAAC--T-CA--T-GGC------G    431

POL_CONSENSUS ?aatgta?aaaTagaa?T?g?a?????????gg???????Aaaa??gtaa?aGg?aCagTatTggtaGGa    372
A_U455        TC-GA--CTT------A-TTGT.........--A.....----AGACT-T---T---------------    399
B_HXB2R       TC-GA--CTC------A-CTGT.........--ACAT...---...-CT-T---T--------A------    402
D_ELI         TC-AA--CCC------A-CTGT.........--ACAG...---...-CT-T---T--------A------    399
O_ANT70C      T-----GAC-G-----A-A-A-.........--A.....-GGGAG---CAG--A----------G---    390
O_MVP5180     C-----GAC-G-----G-ACA-.........--A.....--GGAA---CAG--A----------G---    390
CPZGAB        T------C-C-----GA-A-A-.........--GAGA...---...---GT---G--------A------    396
CPZANT        C--A---CCC---C--A-T-G-.........-AC.....-GG-CA---CT--CA------C--T-G---    402
A_ROD         A------G--------G-TCT-.........AAT.....----AG---CGG-CC--CA--A--AC---C    489
B_EHOA        A------G--------G-A-T-.........--A.....----GA----G--CA------A--AC---G    546
SD_MM251      A------A--------G-TTT-.........--C.....----GGA-T-A---G---A-CA--AC---G    549
STM_STM       G------A-C-----GG-ATT-.........--C.....--G-AAA---A---A--TA-TA--ACT---    504
VER_AGM3      TG--AGGG--G--AG-T-G-A-.........-AC.....----TTT-G-G---G--CA-----A-----    519
GRI_AGM677    GGGG---C--G--C--T-G-AG.........-AT.....----TCA-C-CC--CT--A-TC-AA-----    513
SAB_SAB1C     C---CAGG-GG--A--A-A-A-.........-AT.....----CTTGT-A--CT--CA-T---------    498
SYK_SYK       A---A-TAG-T-CACTG-T-TTAAACCCTCA--AAAAAGA--GCAA---GAG--G--TC-T--A--G--T    501
```

```
                                                                  prot \/ p66, p51
POL_CONSENSUS cc?Ac?CC?gT?AAcATtaTtGGaAGaAAtaT?tTgacac??ttaGG????acttTAaattT?ccaataa    430
A_U455        --T--A--T--C-----A--------G-----G-----T-AGA-T--TTGT-----------T-----T-    469
B_HXB2R       --T--A--T--C-----A------------C-G-----T-AGA-T--TTGC-----------T--C--T-    472
D_ELI         --T--G--T--C-----A--C---------T-G-----C-AGA-T--CTGC-----------T-----T-    469
O_ANT70C      --T--T--T--T--T-----------------A------GGA-----TTGT--AC----C--C--T----    460
O_MVP5180     --T--T--T--T--T---C----G-----C--A------GGA-----ATGT--AC-------C--T----    460
CPZGAB        --A--A--T--A--T--A--------------T-------AA--G--TTGT---C--GTG--C-----T-    466
CPZANT        --A-AC--A--A--------A--T------G-T--ATGT-TT-----ATGT-----------T-----T-    472
A_ROD         GAC--C--AA-C------T----C--------TC-----GCC-----CATGT-A------C-A---G-CG    559
B_EHOA        GAC--C--AA-A------T----C--------T--A-ATAGC-----CATG---C-------C---G--G    616
SD_MM251      GAC--T--GA-T------T----T--G---T-GC-A---GCTC-G--GATGT--C-----C-T--C---G    619
STM_STM       GAC--C--CA-T------T----C--G--CC-A------GCT--G--GATGT-A--------T---G--G    574
VER_AGM3      AGC--T--CA-A-----A--------------A--AG---CAGC---AGCC-AA---GTAA-GGGTCA-C    589
GRI_AGM677    AGT--A--AA-C--T-----A-----------T--AG-T-AGGC---CATG-AA---GT-A-GGG-G-TC    583
SAB_SAB1C     GAA--A--A--T--T-----A---------G-C--AG---AA-----AGTT---C-------GA--CA--    568
SYK_SYK       --T--G--A--A--T--AT-A-----G-----AC-T--CAAA--G--AGTC-AA---GTAA-GGTTCA--    571

POL_CONSENSUS g??c???????ataga?ccagtaaca???gTaaaatTaAAaccaGgaa?ggatggaCCaaaa?TaAaaCA    485
A_U455        -TC-T......--T--AA-T---C--...----------------A--T------C-----GG-T-----    530
B_HXB2R       -CC-T......--T--GA-T---C--...-----------G-------T------C------G-T-----    533
D_ELI         -TC-T......--T--AA-T---C--...-----------G-------T------C------G-T-----    530
O_ANT70C      -CC-C......----CC-----GC--...------C------------T-------------G-------    521
O_MVP5180     -TC-C......----CC-----GC--...-----GC------------T-------------G-------    521
CPZGAB        -TT-A......--T--AA-T--CC--...--C-----G----------T------T-----GG----G--    527
CPZANT        -TAAA......G-T--AA----GC-C...--T---------GA-----T-----------G-G-------    533
A_ROD         CCAAA......G----G---A---A-...A---TGC----G-----G-AA------------C-G-G---    620
B_EHOA        CAAGG......-----A-------A-...--CC-G-----G--T-A--AA-----G------A-C-G---    677
SD_MM251      CTAAG......G----G--T----A-...--C-CC-----G-------A--T----------T-G--G--    680
STM_STM       CTAAA......G----A-------AG...--G-C--------------AA------------A-------    635
VER_AGM3      TGT-AGAA...CA-ATT--CA-T--CCCT--G---------GA---GGCTAGA-----TTTCT-------    656
GRI_AGM677    TAT-TAGTCAG--T--GGA-AC--A-...---C--C-----GA---G-AA--------T---T-G-----    650
SAB_SAB1C     -AGAG......-----A--TA---A-...--CC-C-----G------CAA-----G----GGA---GG--    629
SYK_SYK       CAG-A......T----G-----C-A-...----GTC-G-------AC-AA--ATTG--CCGGC-------    632

POL_CONSENSUS aTGGCC?tTatCaaaAGAaAAaaTagaaGC?tTaa?aGaaAT?tgtgaagAaaTggaaaaggaaggaaAa    551
A_U455        ------A--GA--G----G------A----A----C------T---A-T--G------------------    600
B_HXB2R       ------A--GA--G-----------A----A---GT------T---AC---G--------------G---    603
D_ELI         ------A--GA--G-----------A----A----C------T---AC---T------------------    600
O_ANT70C      ------CC----T-----------------C--G-C--C---A---C-G---------C-A---------    591
O_MVP5180     ------CC----T-G---G-----------AC---CT-C---A---C-----------C-----------    591
CPZGAB        ------C------GC---------TA----C----C------T---C-------------------G---    597
CPZANT        ------GC-C--------G--------G--CC---A------T-----TA-GT-A---GCA---AAT--G    603
A_ROD         ------C---A-------------------AC---A------C------A----------A-----CC-G    690
B_EHOA        ------CC----C-----G------CT---CC-C-A------C------A----------A--G---C-G    747
SD_MM251       G-----A--------------G----TT--A----G------C------A-G-----------T--TC-G    750
STM_STM       ------A--------------G-----G--AC---A---G--C------A-------G-----T--GC-G    705
VER_AGM3      ------CC-C--C-------------A----C---CAG-----A-----CC--T-A--G--A---------    726
GRI_AGM677    ------C-------G---------T-----T----C------A--CA--C---------G----G------    720
SAB_SAB1C     ------T--G-----------------G--C----AG-CC--T--------CT-A-----AC-----C-T    699
SYK_SYK       ------A-----GGT---G---T----G--CC-C-AG-C---AGTA-----C---CTC----C----C--    702

POL_CONSENSUS aT?tcaaaaatagg?cctgaaAAtcCaTAcAAtaC?CCaat?TTtgc?ATaAagAAaAAaGAcaa?ac?a    614
A_U455        --T--------T--G--------------------T---G-A-----T-----------G----GC--T-    670
B_HXB2R       --T--------T--G--------------------T---G-A-----C-----------------GT--T-    673
D_ELI         --T----G---T--G--------------------T-----A-----C-----------------GT--C-    670
O_ANT70C      --T----G------A-----------T--T-----A--T--C-----T-----A--G-----TGGT--T-    661
O_MVP5180     --C----G------A-----------T--T-----A--T--T-----T-----A--G-----T-GC--T-    661
CPZGAB        --A--G--------A--------------------T--T--T-----A--C--A--G-------GT--T-    667
CPZANT        --T--T-G------G--A--T--C-----------A-----T-----A-----A--G-------CTT-A-    673
A_ROD         C-AGAGG--GC-CCT--AACT-----T--T-----C--C-CA-----A--C--------G-----A-AC-    760
B_EHOA        T-AGA-G-GGCGCCT---ACT------------T-G--C-CC--C--C-----A--G--------A-AC-    817
SD_MM251      T-GGAGG--GCTCCC--GACC-----------C--C--C-CA-----C--------------T--G-AC-    820
STM_STM       C-AGA-G--GCTCCT--AAC------T--------C--C-CT-----T-----------G-----A-AC-    775
VER_AGM3      --TAGC--G-----AGGA--G---G-------C--T---G-G---TGC-----------------GT-AC    796
GRI_AGM677    T-A--T-GG-----AGGA--------T--T-----A---G-G-----C-----------G--T--A--AC    790
SAB_SAB1C     T-GGA--G------G--A-----------T--C--C---G-C-----A----G---G-----T--A--TC    769
SYK_SYK       T-GGAG---GCTA-T---ACC-----T--T-----C---G-A----TA--T-G------G-----A-AG-    772
```

```
POL_CONSENSUS AaTGGAGaAtgcTagTaGAtTTcAGagAacTaAAtaagagaAC?CAaGA?TTtt?gGAagT?CAgtTaGG   680
A_U455        -------G-AAT-------------------C-----A-----T--G--C--C-G------T--A-----   740
B_HXB2R       ---------AAT-------------------T----------T-----C--C-G------T--A-----   743
D_ELI         -G-------AAT-------------------T----------T-----T--C-G------T--A-----   740
O_ANT70C      ---------AAT-G--------T--G---T------------A-----G--C-G---G--A---C----   731
O_MVP5180     -G-------AAT-G-----C----------T-------A-----A-----T--C-G---G--G--A-----   731
CPZGAB        ---------AAT-------C-------------------------A-----C----G------G-------   737
CPZANT        ---------A------T------------T-------A-----A-----T----G---GA-A--A-----   743
A_ROD         -------G------A-------------------C---GT---T-----T--CACA---A-T--------   830
B_EHOA        -------G------A-------------------C---GT---C-----A---ACA--G--C---C-G--   887
SD_MM251      -------------GA-------T--G----------G-GTC--T--G--C---ACA-----C--A-----   890
STM_STM       -G-----G-----GA-------------------AGTG--A-----T---ACA---A-T---C----    845
VER_AGM3      ----------T----------T--G--------C--AGC---A-----T----TC-----A--------   866
GRI_AGM677    -------------T------------G--------C--AGCT--T-----C----TT--G--T---C-G--   860
SAB_SAB1C     -G--------A--CA-G--C-----GC-GT-------A--T--T-----C---CA------C--------   839
SYK_SYK       -------G-----CA-T-----T---A--T-----G--GCC--A-----T----TC--G--A--A--G--   842

POL_CONSENSUS aATaCC?CAccCagcaGG?tTa?aaaaaaagaaa??a?T?ACagTa?TaGAtgTagg?GAtGC?TAtTtt   740
A_U455        ---C--G--TA----G--TC--A----G--A---TC-G-A------C-------G--G--C--A------   810
B_HXB2R       ------A--T--C-----G---A----G--A---TC-G-A------C-G-----G--T-----A------   813
D_ELI         ------G--T--T-----GC-GA----G--A---TC-G-A------C-G-----G--T-----A------   810
O_ANT70C      T--C--A--T--G-GG--T--GA-GC-----C--TCTG-T-----CT----------A-----T-----C   801
O_MVP5180     T--T--A--T----GG--T---A-GC---G-C--TCTG-T-----CT----------A-----T-----C   801
CPZGAB        C-----T----------G---A----G--A---TC-G-G------T----------A-----C--C--C   807
CPZANT        ---T--T--T-----C--A---A-GC--------TC-G-G------T-G-----G--A-----C-----C   813
A_ROD         ---T--A----------G--GGCC--G----G-AG-A-T--T---C----------G-----T--C---   900
B_EHOA        T--T--T----------AC-GGC-TC-------AG-A-A------C----------A-----C--C--C   957
SD_MM251      ------A-----T-----AC--GC-----G----AGGA-T------C-G---A----T-----A-----C   960
STM_STM       T-----A-----T-----AC--GC---G-G--G-AG-A-C------T-G--------G-----A------   915
VER_AGM3      C-----T--T---T----G--CG----G-T--CGGA-A-A------T----CA----G-----C----A-   936
GRI_AGM677    ---T--T--------G--CC-TC-G--------GCA-A-C------A----CA----G-----C----A-   930
SAB_SAB1C     G-----C----------C---C-GC---G-G-GCAGA-T-----GT-G--CA----A-----C------   909
SYK_SYK       ------C-------G---AC--...---CG-C--AA-C-A---A-TA-----T--AAA-----A----AC   909

POL_CONSENSUS tca?t?CC?tTagat?aagacTTtagaaaaTAtACtGC?TTtAC?aT?cCaag?gTaAAtAATgag?c?C   800
A_U455        ---G-T--T------G--AG--------G--------G--C--C--A--T--TA----C------A-A-   880
B_HXB2R       ---G-T--C------G-------C--G--G--------A-----C--A--T--TA----C------A-A-   883
D_ELI         ---G-T--C------G----T-----G--------C--C-----C--AT-T--TA----C------A-A-   880
O_ANT70C      ---TGT--C-----CCC---T----------------T--C--T--T--T--T--G---------A-C-   871
O_MVP5180     ---TGC--T------CC----------------C-----C--C--T--T--T--T--G--C------A-C-   871
CPZGAB        --TTGT--CC-G---A----T--C--G----------A-----A--T--C--TA----------A-C-   877
CPZANT        --CA-A--C------C-G----------------A--T--C--A--T-----C--G--C------A-C-   883
A_ROD         --CA-A--AC--C--G-G---------CC---------A-----TC-A---TCA--G--C----CAGAA-   970
B_EHOA        AGTG-C--AC-----CC------C---C--------A--A-----TT-G---GCA----------CAGAA-  1027
SD_MM251      --CA-A--TC-----G----A-----GC-G--C-----C-----TT-A---TCA----------CAGAG-  1030
STM_STM       --CA-A--CC-----G-G-G------GC-G--------T--C--CT-A---TCA-----C----CAGAA-   985
VER_AGM3      ---A-A--A-----CCC---G--------G-----C--T-----C--T---TCA---------C-AGGG-  1006
GRI_AGM677    ---A-A--A---TGCA-G--A--C----------A--A-----C--C--CTCA---------ACAGGG-  1000
SAB_SAB1C     AGCTGC--T------CC-------CAG--------A--A--C--C--C---TCA--C------AG-GAG-   979
SYK_SYK       ---G-C--CC-G--CA----G-----GCC---C--G--A-----AG-G-----CA---------CCT-C-   979

POL_CONSENSUS CaGGaataaGaTAtcagTacAatgt?cT?CCaCAgGGaTGGAAaGGaTCaCCagCaAT?TTcCAaa??ac   865
A_U455        -----G-C--G----------G--T--G----------T----A-----G-GT-G                 950
B_HXB2R       ----G--T-----------G--T----------A-------GT-G                           953
D_ELI         ----G--T-----------G--T----------G-----A-------GT-G                     950
O_ANT70C      ----------C----------C--C--G--A----------T--------T--A-------GTT-       941
O_MVP5180     -----G-------C----------C--C--G--A--G--------T--------C--A--T--G-GTT-    941
CPZGAB        -----G-T--------A--T-----TT-G-----A----------G--T---T-T--C-------GC-G    947
CPZANT        --------------TGT--T-----T--A-----A--C----------T--T--T---GCA-G          953
A_ROD         ------A------CATA--T--A--CT-G----------G----------T--T---CAC--          1040
B_EHOA        ------AG-------TT-----A--C--A----------G-----C--------T-----GTAC--      1097
SD_MM251      ------A-C----CATT--T--G--T--G--T----------G--G--------C--C------TAC--    1100
STM_STM       ------A------CATC-----G--T--G----------G----------T--T---TAT--          1055
VER_AGM3      ----T-CT----------T---CTGT--T-----A--------G-----C---A-T--T--T--G-AC--  1076
GRI_AGM677    ----G-----G-------T----TGT--G--T----------T--TA----T-----G-AT--         1070
SAB_SAB1C     ----C----------T--G--G--C--------G--------C-----CA----T--T--G-CA--      1049
SYK_SYK       -G---GA---------A-TT-CA--GT-A----------T--CA-T--C-------GC--            1049
```

```
POL_CONSENSUS ?atgacaaa?aT?cTaga?ccatT?Agaaaa?a?aatccaga??Tag?taT?t?tCAaTAcATgGAtGAc   923
A_U455        C---------A--CT----G--C--T---TC-C-AC-------CA---T---C-A---------------  1020
B_HXB2R       C---------A--CT----G--T--T------C-A--------CA---T---C-A---------------T 1023
D_ELI         C---------A--CT----G--C--T------C-A--------AA-G-T---C-A---------------T 1020
O_ANT70C      A---------A--T-----T-----C--G-G-G-C--C-----AT---AA--T-G---G------------  1011
O_MVP5180     A---------G--T-----T-----T------AGC--C-----AG---AA--T-A---G-----A------  1011
CPZGAB        T---------A--T-----A--C--C---G--A-G-----T--CA-TAC---T-AC--G------------  1017
CPZANT        C-----C--G--CT---CA-----T--GG-TA-GT------CAG---AA--T-A---------------T 1023
A_ROD         A----G-C-GG-AT----A-----C------GCA--CAAG--TG-CAT---CAT---G-----------T 1110
B_EHOA        C---G----GG-A-----C--T--C------GCC--CAAT--TG-CAC---AATC--G------------  1167
SD_MM251      T----G-C-TG-G-----A--C--C--G--GGCA--------TG-GACCT-AGTC--G--T---------  1170
STM_STM       A----G---T--AT----G-----C--G-G-GCC--C-----TG--AC-C-GATC--------------T 1125
VER_AGM3      AGCAG-TTCC--T-----AGA-A-A-A---GG-GTTAAA-CCCC--ACC--TGTG--------------  1146
GRI_AGM677    GGCAG----C--TT----GGAGA-C-A--GGC-C-C---T-GGT---AA--TGTC-----------C--T 1140
SAB_SAB1C     AGCC-AC--A--T--GC-GGA---T--GC--A-G--C-----TG---A---A-A--------------T 1119
SYK_SYK       A--A-ATC-G--A---C-G-----T-----GA-GT--T----TC-CAC-C-AATA--G----------- 1119

POL_CONSENSUS ?T?t??gTaGg?tctgA??????aga??t?g??catcataga??a???gTagaa?agcTaagagaacatc  971
A_U455        T-G-AT--G--A-----T...TT---AA-A-GG--A------GC-AAAA----GG-AT------CT---- 1087
B_HXB2R       T-G-AT-----A-----C...TT---AA-A-GG--G------AC-AAAA----GG----G---C------ 1090
D_ELI         T-G-AT-----A-----C...TT---AA-A-GG--G-----GAC-AAAA----GA-AT----------- 1087
O_ANT70C      C-A-AT-----A--A--T...TT-CCCC-GACAG-A------AA-AGGA-T---TT---T--------C- 1078
O_MVP5180     T-A-AT-----A--A--T...TT-CCAT-G-CAG-A------AAGAGG--C---TT---T--G------T 1078
CPZGAB        C-A-AT--G--G-----T...CTT--AA-T-AT--A------AA-AAG--G---G-A------C-----T 1084
CPZANT        C-C-AT-----A-----T...ATG--AA-TACTGCA------GA-ATGA-----A----T---C-----T 1090
A_ROD         A-C-TAA---CTAG---CAGGAC---TT-A-AA---G----GGT-...--CCTGC----C-AG----T-- 1177
B_EHOA        A-TCTC--G-CAAG---CAGGAGC--TC-G-AG---G-C--GGT-...--GTCTC-A----A---G-TAT 1234
SD_MM251       A-C-TAA---CTAG---CAGGAC---CC-G-AA---G-C--GGT-...--TTT-C------AG----TCT  1237
STM_STM       A-C-TGA---CCAG---CAGAAC---TC-A-AG---G-C--GGT-...--TTT-C--T-G-AG----TC- 1192
VER_AGM3      C-A-GG-----G---C-GGAAGAT--ATACACG---G--C-GTTG...------C-A------ATGA-AT 1213
GRI_AGM677    T-G-GGT-G-CG--A--CCATGAT--GACTAGA---A--CA-CAG...-----CATAG-----A-GATG- 1207
SAB_SAB1C     A-G-TAA---CCAG---TAGGCC-A-GGCA-AA----TAGT-ATG...---C-GC--T-------CT--- 1186
SYK_SYK       T-A-TGA-T--CA----TAGATC---GAAG-CA-----GGAGAT-...---C--C-AA--GTGAC-GCA- 1186

POL_CONSENSUS Ta?ata??tggGGatT?ac?accCCagA?aAaAAa??tCAgaaagAacctCCatttcaaTGGaTgGG?TA 1032
A_U455        --TTG-GC--------C-TT--------C-----GCA--------------------TT--------G-- 1157
B_HXB2R       -GTTG-GG------C-T--C--A-----C------CA-----------------C-TT--------T-- 1160
D_ELI         --TTG-GG--------T--C-GA-----T------CA-----------C-------TT--------T-- 1157
O_ANT70C      --T--CAG--------C--T-----T--C-----GCA---A--G--------C----T---------G-- 1148
O_MVP5180     --T--CAG--------T--T-----T--T-----GCA------G--------C---TT---------A-- 1148
CPZGAB        -GCT--AA-----G--C--A--------C------CA---A--G--G--A--C---TT---------A-- 1154
CPZANT        --C-GGTC------C-AGAG--T--T--C-----GCA-------------------------------A-- 1160
A_ROD         --A--GGCCTA-----TT-T--------TG-G--GTTC--A-----C-------AC--C--------C-- 1247
B_EHOA        --A---ACAT------CT-T--T-----AG----GTTC--A-----C--------CA----------G-- 1304
SD_MM251      --A---GCATA--G--CT-T--------AG-G---TTC--A-----T--C-----------------G-- 1307
STM_STM       --A---AT-TA-----TT-C-----T--AG-G---TT------G--T--------------------A-- 1262
VER_AGM3      --AG-GCC--------AGAA--A-----C--G---GTA------A----A--T-A-G-G--------A-- 1283
GRI_AGM677    -GCTAGAAAAA--TC-AGAA--------C--G---GTC--A-G------G----GGG----------G-- 1277
SAB_SAB1C     --G-A-CT-----G--T-AA-----T--A--G--GTT---A--G--T--A----AC-TC--------G-- 1256
SYK_SYK       --TTA-AAGT------T-AGGT----A-AG-----TGG---G-CC--TA----A-G--G---T-A--T-- 1256

POL_CONSENSUS tgaacT?catCCaga?AAaTGGaca?T?cagaa?ATa?aatT?ccagaa???aaaGA?g?????tggACa 1086
A_U455        ------T-----T--C---------G-T---CCT---C-GC-G------...-----CAGC...-----T 1221
B_HXB2R       ------C-----T--T---------G-A---CCT---GTGC-G------...-----CAGC...-----T 1224
D_ELI         ------C-----T--T---------G-A---TCT---A--C-G------...--G--GAGC...-----T 1221
O_ANT70C      ---G--C---------C---------G-A---TCC--CC----G--TA-C...--G--T-TG...------ 1212
O_MVP5180     ---G--C--C-----C--G------G-A---CCC--CC----G--T--C...-----A-TG...------ 1212
CPZGAB        ---G--C--C-----C---------G-C---CCT--TC----A------...-----G-TA...-----T 1218
CPZANT        ---GT-A---------C---------TG-A-----A---A-GC-A-----G...CC---T-AT...----- 1224
A_ROD         ------ATGG---ACT-------AGT-G-----A---C-G--G--CC--...-----AATA...------ 1311
B_EHOA        ---G--CTGG---A-G-------A-C-G--A--A---C-GC-A------...-----G-TT...------ 1368
SD_MM251      C---T-GTGG--GACA-------AGT-G--A--G---G-G--G---C--...-G---GACC...------ 1371
STM_STM       ---GT-ATGG---ACA--G----A-C-G--A--A---G----A--TC-G...-G---C-TT...------ 1326
VER_AGM3      CA--T-GTGG---C-C--G---CAGA-AAGC-GC---G----AGA---C...-----A-AA...-----T 1347
GRI_AGM677    -A--T-G-----GA-T--------CA-TA-C--A---G----A--CCCC...TT---A-GAGAA------ 1344
SAB_SAB1C     ----T-GT-----A-G------CA-C-G---G-G--CACCC-A------...-GG--A-AA...------ 1320
SYK_SYK       CACT--T--------C--G---CAGT-G-----A---G--C-C--TA-TATAG-T--C-AA...ATC--- 1323
```

```
POL_CONSENSUS GT?AAtgAtaTaCAgAaatTagTaGgaaaatTaAAtTGGGCAagtCAaaTtTatccaGGaaTtAaaac?A  1154
A_U455        --C---------------------------C---------------G------G----G------GTA-  1291
B_HXB2R       --C-----C---------G-----G--G-----G-------------G-----C-----G------GTA-  1294
D_ELI         --C-------------C-----G-AG-G------C---------C--G-----------------GTA-  1291
O_ANT70C      --A-----------A---C--A-------GC-------------------C----A--------G-GTG-  1282
O_MVP5180     --A-----------A-------------------------------------C----A--------G-GTA-  1282
CPZGAB        --C-----------A---C-GA-------G-----------------G---------------A----TA-  1288
CPZANT        --T-----C--C------C-------------------------G--C--C--------C-----T-     1294
A_ROD         --C-----C--C-----GC----G--TGTCC-----------GCA---C-C--C-----G--A--G--C-  1381
B_EHOA        --A-----C--T-----G-----G---GT-------------GC----C---TC--G--G-----G--C-  1438
SD_MM251      --G--------------G---------GT-------------GC---------------T--A-----C-  1441
STM_STM       --G---------C--A---C-G-----GGTC-----C------GCA----------------------T-  1396
VER_AGM3      --A-----------A-G-C----G--G---C-----------GCA--GC----C------C-C-G---T-  1417
GRI_AGM677    --A--CA-A--------GG--------GTTC-------------------------------------C-  1414
SAB_SAB1C     --G--------T--A-------------G----------------------CA----G--A--G--T-  1390
SYK_SYK       --G---C-GT----------GA-----GT-C-T----------C--G------T----------G--A-  1393

POL_CONSENSUS aacA?tt?TGtaaa?taaT?AGaGGaac?Aaaa?????cTaacaGAagtagTacc?ttgAcagaagaaGC  1213
A_U455        -G--AC-G------C-TC-C--G---G-C---GCA...--------TA-----A-CC----T--------  1358
B_HXB2R       GG--A--A------C-CC-T--------C---GCA...-----------A----AC-A-----------  1361
D_ELI         G---A--A------C-CC-T--G-----C---GCA...-----------A----AC-A-----------  1358
O_ANT70C      G-G-A--G-----GT----T-----C--C--GTCA...T----------------T--A-GTAG---G--  1349
O_MVP5180     --G-A--G--C--GT----C--------C---TCA...T-G-----G--------T--A-GTA----G--  1349
CPZGAB        -G--A--A-----GC-G--A--------A--G-AA...--G-----T-----T--TC-C---CC------  1355
CPZANT        -G--G--G----G-C-C--C------GTC--G-GT...---------TAG-----AAA----TAGG-----  1361
A_ROD         ----C--A----GGT----C------...----TGACA--C-------A-----AG-G-------TT---  1448
B_EHOA        GG--TA-A------C----A--G---...--G-TGACC----------AG----AA-G---T---TTG--  1505
SD_MM251      ----TC-C----GGT----T------...----TGACT--------G-A---T-AG-G---T--GATG--  1508
STM_STM       ----TC-G--C---T----C-----T...--G-TGGCA--G--T--G-AG----AG-G-------ATG--  1463
VER_AGM3      --A-C--G------T----C-------AA--G-AC...T--CT----AC----A-C-G------G-----  1484
GRI_AGM677    ----TACC---GCCA-GT-G-----G-AA--G-AC...--CCT-----A-A--GTA-G---------G--  1481
SAB_SAB1C     ----T--G----G-C----T------G-T-G-CCT...T-G-----GA----C-AA-G---------G--  1457
SYK_SYK       --G-GC-C-----GTGT--C--------T---CAA...T----------C-CA-AC-T-------C---  1460

POL_CONSENSUS aGAa?taGAatt?gaagAaAAcagaga?At??????????cTaaaa?a?gaacaagaa???GGa???TAt  1262
A_U455        ----T-------G-C---G-----G--G--T.........------...--C-CT-TGCAT---GTA---  1416
B_HXB2R       ---GC-----C-G-C----------G--T.........------...----C--T-CAT---GTG---  1419
D_ELI         ----T-----C-G-C---------G--A--T.........T------...----C--T-CAT---GTG---  1416
O_ANT70C      ---GC-G-----A--G----------A-GG.........T------...C---C--TGCAT--GGTA--C  1407
O_MVP5180     ----C-------A--------------A-AG.........------...--G-C--T-CAT---GTA---  1407
CPZGAB        ----T-------A-C------T--G--G--A.........G---GC...AC--C--T-CAT--GGTA--C  1413
CPZANT        ----T-------A--------T-A-C-A--TNNNNNNNNN---C-GC-GA--AT---G...---TAC--C  1428
A_ROD         ----GC---GC-A----------ATT--C.........----GCC-G--------G...---CAC---  1506
B_EHOA        ---GGC------CC-G-------A-ATC--C.........---G--C-A--G--G---...---TCC---  1563
SD_MM251      ----GC-----AT--G-----T-AGATA--T.........--C-GTC-G---------...---TGT---  1566
STM_STM       ---GGC-----AT-C------T-A-ATT--T.........--C-GTC-G--G--G---...--GCAA--C  1521
VER_AGM3      ----GC-----AT-C--------A---G--C.........T-----ACG-----G---...--GACC--C  1542
GRI_AGM677    ---GGC-----ATA-GA-C--TCA--GG--A.........G-GC-GG-AAC-------...---ACA--C  1539
SAB_SAB1C     ----C-------A--------T---C-G--C.........T---G-C-AA-G--GC-G...---CAA---  1515
SYK_SYK       T--GGC------A-----G--T---C-A--C.........T----GG-A-----G-C-...--GTCT---  1518

POL_CONSENSUS TAt?a?ccag??aaaga?tTa??aGcaa?aaT?cagAAaca?gg?ga?gg?CAaTGGacaTAtcaaaT?t  1318
A_U455        ---G-C---TCA-----CC--GT----GA---A--------A--GC-A-AC----------------T-  1486
B_HXB2R       ---G-C---TCA-----C---AT----GA---A-----G--G--GC-A--C----------------T-  1489
D_ELI         ---G-C---TCA-----C---AT----GA---A--------A--GC-C--C-----------C-----T-  1486
O_ANT70C      ---C-A--T-AT--G--TC--TGG-TT-AT--T-----G--A--A-GG-AG--------T--C--G--A-  1477
O_MVP5180     --CC-G--T-AC-----C--GTGG-TT-GT--T-----G--T--A--A--G--------T--C--GG-A-  1477
CPZGAB        ---G-T----AC-----GC-TAT----GA---A--------G--CA-CT-C--G-----T-----G--A-  1483
CPZANT        ---C-G--T-GTCT-CCAC--AA-----CC--A--------A--GTCA--A--------C--------A-  1498
A_ROD         --CC-AGA--AA-----GC--GA-----C-G-C--A--GG-TCAA--GAAT--G---------A----AC  1576
B_EHOA        --CA-AGA--GGGT-CCT---GA-----C-G-G------A-TCTA-CAAAT--G---------CA-G--TC  1633
SD_MM251      --CC-AGA--GC--GCCA---GA---C-CGG-AATA--GAGTCAG--CAAT--G---T-T---A----TC  1636
STM_STM       --CCGAGA--AC--GCCTC--GAG----C-G-AGT----G-TCAG--CAAT------T-C---A-G--AC  1591
VER_AGM3      --CA-A----GA-G-CCCA-CAG----GC-G-G--A----TA-AA-GA--T-------GT--C---T-CA  1612
GRI_AGM677    ---G-C--TCTC-----A---AT-----C-G-T--A--G--A--A--A--G----------C---T-CA  1609
SAB_SAB1C     ---G-C--T-CTCTTCCT---AGG--T-AGG-T-TT----TG--A--T--C------GG------G--A-  1585
SYK_SYK       ---G-C--TAAA---CCT---GA----CAC--CACT----TA--GTCCCAG------GG---CATG--TA  1588
```

```
POL_CONSENSUS .atCAagaagaa??????aaaat?cT?AaagcaGGaAaatatgCaAag???aa?????C?CAcAc?AATGa 1370
A_U455        -------GCC-TTT...----AT--G---A--------------GAAAA-GGTCTG-T-----T----- 1553
B_HXB2R       -------GCC-TTT...----AT--G---A--------------GAATG-GGGGTG-C-----T----- 1556
D_ELI         --------CC-TTT...----AT--G---A-------G-------GAATG-GGGGTG-C-----T----- 1553
O_ANT70C      ----G------CAT...--G-AC--C---A----G------A-T-G-CAA--GGCCT-C-----A----- 1544
O_MVP5180     ----G--T---CAT...--G-AC--T---A----------T-G-CAA--GGCCT-C-----A----- 1544
CPZGAB        T---G---CC-CAT...--G-ATT-G--GA----G--------C-GACAA-GGTCAG-A-----A----- 1550
CPZANT        -----A-T---GGA...---C-GT-A----------------T-G-CCT-CAGGAA-T-----T----- 1565
A_ROD         -C--G------.....-----T--A----T----------GTG--AAACA-C--T--C----G 1640
B_EHOA        ----G-G---T.....-----C--A----T----------GTT--AAACA-T-----C----G 1697
SD_MM251      -C--------C.....-----A--G----T--------T-------ATA--GAATA-A--T--C----G 1700
STM_STM       -C--------C.....-----AT-G----T---C--G-T------AATA--AAACA-A--T--A----G 1655
VER_AGM3      -G-----G-G-.....C--G-AT-A----T---T-----CA-----CAG--AAACA-T--T--C----- 1676
GRI_AGM677    CC-------GG.....GC-G-AT-A--G-TG----G------C---CAA-GAGAAA-T--T--T----- 1673
SAB_SAB1C     -C--GCC----AAT...-----A--G--G-TT--C----------AATC--AACAG-A-----C----- 1652
SYK_SYK       -G-----GC--AAAGGCCC-CCC--A-T-A-------GACA-----AACCTTTGCAG-A--T-GT----- 1658

POL_CONSENSUS ??taagac?atTagCaga?gtagT?cAaAaaaTagg?aaagAA?c?aTagt?at?TGGGGaaagat?CCa 1429
A_U455        TG---A--A----A----A--G--G------G-GTCC-C----AGC-----A--A----------C--T 1623
B_HXB2R       TG---A--A----A----G-C---G---------ACC-C----AGC-----A--A----------CT--T 1626
D_ELI         TG---AG-A---------G-C---G----G----TCC-C----AGC-----G--A-------G--CT--T 1623
O_ANT70C      TA------A---------A---A-C--G--GG-GTCTC-----T-T---A-T--C--------AT-G--T 1614
O_MVP5180     TA------A---G-----A-----C--G--GG-GTCTC-----G-T-----T--A-----G--AT-A--T 1614
CPZGAB        CA-C--G-A-C-G--G--A-C---A--------T-CT-CT---AGC-----A--T--------A-CA--T 1620
CPZANT        GG-T--G-A------T-GA-----G----------ATT----AGT-----T--C------G--G-A--- 1635
A_ROD         AA-C---TTG------C-G-----T--G--------A------G-AC----C--T------CGA--A--- 1710
B_EHOA        AG------T---G--TC-T-----C----------A--G---G-AT-G--C--C------G----A--- 1767
SD_MM251      AG-T----T-------C-T---A-A--G--------A--G---G-A-----G--C------C--G-C--- 1770
STM_STM       AG-T---TT---G---C-T--GA-T--G--G-----G------G-A-----G--C------C----C--- 1725
VER_AGM3      GT-CC-TGT---G----GAT----A------C-TT-T------T-TT----T--A------G--T-G--- 1746
GRI_AGM677    TC----GACTC-----C-CC-T--C--------CT-T--G---G-AC-TACC--T-------GAC-T--- 1743
SAB_SAB1C     AT-----ATG-------GCC----A--G--G-----A--G---AGC-----C--C------C----A--T 1722
SYK_SYK       CTATCA-TCTC-T--TC-AT-GT-AA-T--------G-TCC--T-AC--TGGTAT---------G-A--- 1728

POL_CONSENSUS aaatTt?a?cT?CCagTa?a?AgagAaacaTGGGAa??aTGGTGG?cagA?TAtTGGCAgGc?accTGGa 1489
A_U455        ------AGA--A--CA--C-A-AG----------GC-------ATG--G----------T------- 1693
B_HXB2R       ------A-A--G--CA--C-A-AG----------AC-------A----G--------A--C------- 1696
D_ELI         ------AGA--A--CA--C-A-AG----------AC-------G----G--------A--C--T---- 1693
O_ANT70C      ------A-G--G-----CACT--------T------AC-------G-G--C--------A--C------- 1684
O_MVP5180     -----CAGG--G-----TACT--------T------ACT------G----A----------C------- 1684
CPZGAB        ------AGGT-A------C-G-A-----GT------GC-------G----A--C--------A------- 1690
CPZANT        ------C-AT-A--CA--ACT--G-----T-----TGCC------T----C----------C------- 1705
A_ROD         ------C-C--A------G-G-------TC-----GCAG------GATA-C--C-----A-TG--A---- 1780
B_EHOA        -TG--CC-T--A------G-A-----G---------TCAG------A----T--C-----A-TA------- 1837
SD_MM251      -----CC-CT-A-----TG-G--G--TGT-------CAG------A----C----------TA------- 1840
STM_STM       -----CC-CT-G------G-G-----GGTC------CA-------G----T----------TA--A---- 1795
VER_AGM3      GTCC--G-A--C---A--G-G--G---GT-------CA-------G-T--T--C--------TA-GT---- 1816
GRI_AGM677    CG-G-AC-A--C------G-C-AGA----------TATG------CAG--C---------A-TAT------ 1813
SAB_SAB1C     -TTA-GG-A--G------G-A-----GCTC------CA-------T-T--C----------TC------- 1792
SYK_SYK       -C---CC-T--C------A-A-----GGA-------AA-------A----C--------A--C--A---G 1798

                                        <- CRS
POL_CONSENSUS T?CCaGAatgggA?tTtgTca?tACcCC?CcatTagT?AaattatggTacaa?tTagtaaaaGAaCCcaT 1553
A_U455        -T--T--------G-------A------T--CC----A----------C-G----AG-----C----- 1763
B_HXB2R       -T--T--G-----G-----T-A------T--C-----G----------C-G----AG---------- 1766
D_ELI         -T--T--G-----A-------A------T--T-----A----------C-G----AG--G--------- 1763
O_ANT70C      -T----------A-------GC--A--C-----GA-C----------GGC---A--GT-----T-- 1754
O_MVP5180     -T--T--------A-------GC--A--C-----GA-C----------C-G----A--C------T-- 1754
CPZGAB        -C--T--------A---A---A------A--C-----C----------GT----AG-C------T-- 1760
CPZANT        -A----------G-----T-A---A--C------A-T-G-C-C-----T--TC--T-GGC---C--T-- 1775
A_ROD         -C-----C-----C--C--GTC------A---C-G--C-GG---GC--TT--CC-G---GGG--T--T-- 1850
B_EHOA        -C----------T------TCA-----A------A-A-GG---GCC--T--CC-G--C-----C---C- 1907
SD_MM251      -A--G--G-----T---A--TCA--G--A---C----A-G----GTC-T---TC----G--G--C--T-- 1910
STM_STM       -A----------T-----ATCC--A--T--T-----A-G----GTC-TT--T---------------T- 1865
VER_AGM3      -T-----C-----A-------G------A--CC----A----------T-CCC-GAC----------- 1886
GRI_AGM677    -A----------G-----T-GC--A--A-TCC----A---C-G-----TTCC-------------A-- 1883
SAB_SAB1C     -C--G--------AA-G----G---T--T-A---GA-T-G---G--------A----------T----- 1862
SYK_SYK       -A------GTCA-A--CA-A-G------G-----G--T-GG-GG-AC--T--C--G---CC--------- 1868
```

```
POL_CONSENSUS a??aggagcaGaaaccTttTAtgtaGATGGagCagctaatAgaga?ac?aaa??aGGaAAaGCaGGaTAt  1617
A_U455        -GC---------G--A--C----------G----------G--A--T--GCT----------G---     1833
B_HXB2R       -GT---------------C----------G--------C--G--G--T---TT----------------  1836
D_ELI         -AT----------T--C----------G----------G--T---TT----------------------  1833
O_ANT70C      TATG--G----------A-------------------------G--A---CT------G---------  1824
O_MVP5180     TGT---G-------------------------------GA-T--A---CT------G--G------  1824
CPZGAB        -CC-ACCA----T--T-A------------------A-----G--A--A---ACT--G----------  1830
CPZANT        TCC--AG--T--------------T-----G----------A-CT-TC--TTG--G--G-----C---  1845
A_ROD         -CC---T-----G-----C--CAC-------T-CTGC-----GC-AT-A---GA--------------  1920
B_EHOA        -GA-----T------T-AC--CAC-------T-CTG---C-A--CCT-A---GA---G----------  1977
SD_MM251      -GAG----A--------A----AC-------T--TG-----A-C-GT-A---GA---G----------  1980
STM_STM       GGA----A-------A--------G-----CT--TG---C--GC-AT-T---GA---G--------C--C 1935
VER_AGM3      -CC-AAG-A---TGT--AC-----G--------TTG-------A-TT-A-GGGA--------------  1956
GRI_AGM677    CAA-----A---TGTT-A------G-----G-----ATCC-A--TG--C---TT---T--G-------  1953
SAB_SAB1C     CCC-----A--C-GT--A----------G----------A-TT-C---GA--------------C    1932
SYK_SYK       -CC--A-----T------C--C----------------------C-GC--GAC------T---------  1938

POL_CONSENSUS gT?aCaga?agaGGaaaaca?AaagT?a?a???tTagaa?A?AC?AC?AAtCAacAagCaGAatTa?a?G  1673
A_U455        --C--T--C-------G---A--G--TGTTTCCC--ACTG-G--A--A------A-GA-T------C-T-  1903
B_HXB2R       --T--TA-T-------G---A-----TGTCACCC--ACTG-C--A--A-----GA-GA-T--G---C-A-  1906
D_ELI         --T--T--C-------G---G-----TGTCCCT--GACTG-C--G--A-----GA-GA-T--G---C-A-  1903
O_ANT70C      --T-----ACA---G-----G--GA-A-T-AAA-----TG-G--C--C------A----T------ATG-  1894
O_MVP5180     --T-----ACA---------G--CA-A-T-AAG------G-G--A--C------A-G--T------ATG-  1894
CPZGAB        --A-----C-A---G-----A---A-C-TTAGC------A-T--C--T-----G----------A-G-  1900
CPZANT        --G-----C-------G-AGT-GG--A-A-CACC--C--A-G--C--C-----------------C-G-  1915
A_ROD         --A-----T-----G---G-C--G--A-AGAAAC----GC-A--T--C-----G---------C--G-A-  1990
B_EHOA        --C-----C--G-----GG-T-----T-A-CCA------C-A--A--A-----G--------GC-TG-A-  2047
SD_MM251      A-C-----T--G--C---G-C-----A-A-GTG------C-G--T--T-----------------G-A-  2050
STM_STM       A-A--T--T---------A-C-----A-A-GCT------C-G--T--T-----------------G-A-  2005
VER_AGM3      A-C---C-ATAT--G-----A-GG--GGA-AAA------A-T--A--A--C--G----------ATG-  2026
GRI_AGM677    C-GT----G---------AGT-G-A-T-GGGAA------A-C--C--T--C--------------ACA-  2023
SAB_SAB1C     T-A-----T-----GG-T--A--G--GGT-GCA-----GA-T--T--C--C--GA-------G---G-G-  2002
SYK_SYK       --AG--AGTGAT--G-C---G-G---ACAGTATC-G--GC-A--A--A--C-----G---------G-G-  2008

POL_CONSENSUS c??Tt???aTgGC?tTgcagGAttc?gga??aaaagtaAAtATagTaacaGAtTCaCAaTAtgcattggg  1734
A_U455        -AA-CCATC-A--C----------A---TC-G-------------------C-----G--------A--  1973
B_HXB2R       -AA--TATC-A--T----------G---TT-G-------C----------C--------------A--  1976
D_ELI         -AA--AATC-A--C----------G---TT-G-------C--------------------------A--  1973
O_ANT70C      -GA-ATTAC-A--C--A--------CAA-GA--C-------------------------------------  1964
O_MVP5180     -TG-ATTA--A--C----------CAAGGAGC-------C----------C----------T------  1964
CPZGAB        -TT-GCTTC----C----------A-ATCA-C-G--T--C-----G--T--C----------TG--A--  1970
CPZANT        -GA--CTT-----TC-AG-------AAC-GGCCC---C--------C---------------------  1985
A_ROD         -CT--GCG-----AC-AACA--C--G--TCC------T-----TA--GT---C-----G----T-A----  2060
B_EHOA        -AT--GCAC-A--AC-A-----C--A---CC-C-G--C-----CA--GT--------------TCA----  2117
SD_MM251      -AT--CTC-----A---ACA--C--A--GCC---GAC------TA--GT--------------TTA----  2120
STM_STM       -AT--GCC-----C---GC---C--A--TCC-----C------T--GGT---C----------TGA----  2075
VER_AGM3      -CA-AAAA-----AC-AG-A---AGT--GCCT--T-----C-------------------------A----  2096
GRI_AGM677    -AG--AAG-----A---G----CAGT---GA---T----------C--------T-------T-A--AA  2093
SAB_SAB1C     -CA--CTGT-A--C--AAGA--C--T---AGC------------A-------C----------TA----  2072
SYK_SYK       GCT-GCTC-----A--------C--AAA-GAC--------------GT----------CT-T-AT--  2078

              CRS ->
POL_CONSENSUS ?aTcat??cagcacaaCCaaca?a?agtgAatca????TAgT?aa?cAaATaATagaagaa?taataaAa  1792
A_U455        G-----TCAG---------GACAGG---------GAAA----C--T----------GA-GC-----G--  2043
B_HXB2R       A-----TCA----------GATC-A---------GAGT----C--T----------GC-GT--------  2046
D_ELI         A-----TCA----------GATA-G---------GAGT----C--T----------GC-GT--------  2043
O_ANT70C      CG----CT-CT-C-----T---C-G--------CCCTA----TC-G----------G---C---C----  2034
O_MVP5180     C-----AT-CT-C----------C-G-----C--CCCTA----TC-G--G--------G---C---C----  2034
CPZGAB        G--T--TCAGT----G---GATC-C---------GAAT----C--T--G----------GT----T---  2040
CPZANT        AG--T-GCA--GTACC---GATC-A---------CCCC----GG-AG-------CC-GA--T--------  2055
A_ROD         G----GTG--AGC---------G-GTCA---AGTAAAA----G--C--G--C---------A-G------  2130
B_EHOA        A--AG-AG-T------------G-A-CA------CCGA----A-GAG-------T------A-G--C---  2187
SD_MM251      A--A--AA---G-TGC--T---G-ATCA--GAGCAGGC----T--C--------------A-G--T---  2190
STM_STM       A-----AA---G---G--C---G-ATCG--GAGCAAGT----A--T--G----------A-G--T---  2145
VER_AGM3      A--AT-AA-T--C-----C---C-G-----C---CCCT--A-AG-A-----T----C-CT-A-GG--C--  2166
GRI_AGM677    C---T-GA-----TGT---CAGG-A---A-C---CCCT----GG-A--G------C---CCC----G---  2163
SAB_SAB1C     T--T--AG---GGG----T---G-ATCA--TAATAACA----AC-A-----T-----G---C-------G  2142
SYK_SYK       C--TC-TATGA--TGC--C--TA-C-CA---CACCCAA----GG-G--G------C----GGCC------  2148
```

```
POL_CONSENSUS Aaggaa?aaaT?TAt?T??caTGGGTaCCaGC?cAcAAaGG?aTaGGaGGaAAtgaagaagTaGAtaAat  1855
A_U455        ------A--G-C--CC-GT-------------G--------G--T-----------GC------------  2113
B_HXB2R       ------A-GG-C---C-GG-------------A--------A--T-----------C------------  2116
D_ELI         ------A-GG-T--CC-GG-------------A--------A--T-----------C------------  2113
O_ANT70C      ------C-GG-G---C-TA-------T--T--C--T-----C---------------A--A---------  2104
O_MVP5180     ------CG-G-G---C-TA-------T--T--T--------C---------------A--A---------  2104
CPZGAB        --A---A----C--CC-CT-C-----------A--------T--------------GC----G-------  2110
CPZANT        -GA---C-G--T--CC-CT-C-----------A--T-----C--------------C-G-----C----  2125
A_ROD         ------GC---C---G-TG-------C-----C--------C-----G-----CC-G---------C-T-  2200
B_EHOA        ------A----A---G-AGG------------T-----G--AC-G--T--T---C-G---------CC-CC  2257
SD_MM251      ---TC-G----T---G-AG-------------A--------T-----------CC-----A----CC-CC  2260
STM_STM       ------GC---T---G-AG-------------A--T-----A----------C-G--G------C-T-  2215
VER_AGM3      --AC-TC-G--A--CT-GCA-----------AG-------G--------C--------GA--------C  2236
GRI_AGM677    ---AGGC-GG-C--CT-ACA-----------T--T--G--G--------C---AC----A---------  2233
SAB_SAB1C     -----GGC-G-G---A-AG----------T--C--T-----AG----T--C-----G---A-T------C  2212
SYK_SYK       ------GCC--A---G-GA-------C-----C--T-----A----------------CT--T------C  2218

POL_CONSENSUS TaGT?AGtcaaGgaAT?AGaaaagT?cTaTTccTaGAaagaATaga??aaGC?CAaGAaGA?CATGaaAa  1918
A_U455        ----C---TCT-----C--G--G--G-----TT----TG-G-----TA-G--T-----G--C--------  2183
B_HXB2R       ----C---GCT-----C--G-----A-----TT----TG-------TA-G--C-----T--A-----G--  2186
D_ELI         ----C-----------C--G-----A-----TT-G--TG-------TA-G--T--------A-----G--  2183
O_ANT70C      ----A--CA-G-AT--T----G---C--G---------G-------CC-G--A--------T--------  2174
O_MVP5180     ----A--CA---AC--T----G---C--G-----G---G-------TC-G--A--------T--------  2174
CPZGAB        ----C---GC------C--G-----G----------TG-------CAGG--C--------A-------G  2180
CPZANT        ----C--------G--C---C----C--C--T--G---G-C-----TA----T-----G--C-----T--  2195
A_ROD         ----G-----G--T--C---C----GT-G-----G----A------GCCC--T--G-----A--------  2270
B_EHOA        ----G--C---------T---C--A-C-----T-------A------ACC---T--------A--------  2327
SD_MM251      ----T-----G--G--T---C----T--C----T-G----AG-----GCC---A--------A-----T--  2330
STM_STM       -G--A-----G-----T---C-G--C-------------A------ACC---T--------G--------  2285
VER_AGM3      ----A--------G--G--G---A-TT----TT------A------AG----C--G--G--A-------G  2306
GRI_AGM677    ----A--CA-------A---C-GA-C--C---T----T--------AG----A-----T--C----C---  2303
SAB_SAB1C     ----T----------A---C----A--------G--C-----T--AG----A--------G-----T--  2282
SYK_SYK       ----C---A-------A------A-C----------------TCCCC----A--G-----T-------G  2288

POL_CONSENSUS aTAtCAtag?AAttggagag?a?Tggc???tga?Tttgg??TaCCacc??taGTaGCaAaagaaATagTa  1977
A_U455        ------CT-C--C-------C-A----TAG---T---AATC-G-----TG-G-----G--G---------  2253
B_HXB2R       ------C--T----------C-A----TAG---T---AACC-G-----TG--------------------  2256
D_ELI         ------C-AC----------C-A----TAG---T---AACC-------CG-G-----------------  2253
O_ANT70C      ---------T----------A--C-C-A--TAG---A-----AC-------AG-G--G--C--G-----CA-T  2244
O_MVP5180     ---------T----------C-T-A--TAG---C-----AT-------AA-------C--G-----CA-T  2244
CPZGAB        G--------T-------A--CTA----TAG---T---AATT-------CA--------------------G  2250
CPZANT        ---------C---------TC-T-A--AGA---A-ACAATC-T--C--TA-T--G--T--------TA--  2265
A_ROD         ---------C---GTA-A--A-C--T-TCA-A-A-----AA-----CAATT----G----GGC--------  2340
B_EHOA        --------AT---GTA-A--A-C-A-TCCA-A-A-----GA-T----AAT----G----G-C--------  2397
SD_MM251      ---C-----T---GTA-A--A-T---TATTCA-A-----AT-----CAGAA----G--C-G-C-G------  2400
STM_STM       ---------C---GT--A--AGC-A-TCTTCA-G-----CA-----AGGC----------GC-------G  2355
VER_AGM3      G--C----AT--------GAACT-A--AGACACT-A---GC------AAA-T--G---------------  2376
GRI_AGM677    G--C----AC---------AGTA---TACAG--A-----AT----TAATA-------------G------  2373
SAB_SAB1C     ---C---GCC--------GAGTA--CAACAA--A-----CT----TG-TA--------------------  2352
SYK_SYK       G--C--CTCA---AT-GAGTATC-TAGGCAG--G---CATC----C-GACAG----------C---TA--  2358

POL_CONSENSUS gc?a??TGTgataAaTGtCA??taAaaGGaGAagC?at?CATGG?CAaGTAgAct??agtccagg?ataT  2036
A_U455        --C-GC---A----------AC-------G-----C--G-----A----------GT--------G----  2323
B_HXB2R       --C-GC--------------GC-------------C--G-----A----------GT--------A----  2326
D_ELI         --T-GC--------------GC-------------C--G-----A----------GT--------A----  2323
O_ANT70C      --T-GC---CC---------TA-------G-----A--T-----T--G-------GC-------AAG---  2314
O_MVP5180     --T-GT---CC------C--TA-------G-----A-CG-----T----------AC--C----AG----  2314
CPZGAB        --CCAT--------G--C--GG-------------C--G-----G--G-------GT--C-----G--T-  2320
CPZANT        --ACAG--------------CG----G--------C-GG-----A----------GC--------A----  2335
A_ROD         AACTCA----CCC-------ACAG-----G-----T--A-----G------A-TGCAGAA-T---C-CT-  2410
B_EHOA        AATTCC----------C--ACA------G-----T--T-----A--G---A-T-CAGAA-T---G-C--  2467
SD_MM251      -AC-CC--------------TCAG-----------T--A-----G--G---A-T-CAGA--T---G-CT-  2470
STM_STM       -AT-CA-----C--G--C--CCAG-----------T--A-----G------A-TGCAGAGTT---G-CT-  2425
VER_AGM3      --C-TG---CCA--------GA-------G---C-AG-C-----G--------TGCCTCG-----GG---  2446
GRI_AGM677    --GGCA---CCC-----C--AA---G-------C-T-AG-----A--G------GCCTCCATT-AA-CT-  2443
SAB_SAB1C     --GGCA---CC------C--GA-------G--GT-TG-A-----G--------TGCT--------AG-G-  2422
SYK_SYK       CAACAA---CCA--------AAAT-G-------C-A-AA-----G--------TGTAGA-AT-TAT-AT-  2428
```

```
POL_CONSENSUS GGCAaaTgGAtTGtACaCAtttaGAaGGaAAagT?aTcataGTaGCaGT?cAtgtagC?AGTGgaTttAT  2103
A_U455        -----T-A-----C------C-------------A---C----------C--------C-----C-AC--  2393
B_HXB2R       -----C-A-------------------------T---C-G--------T--------C-------A---  2396
D_ELI         -----T-A-----------C-------------T---C-G--------T--------C-----C-A---  2393
O_ANT70C      -------A-----C------A-G-----C---A-C--A-----T--T--C-----G--A-----G--C--  2384
O_MVP5180     ----------------------------C---A-C--A-----T--T--C--------A----AC-----  2384
CPZGAB        -----G-A--------C--CC-------C-----G--------G-----T--C-----C-----C-A---  2390
CPZANT        -----G-A-----C--C-----------T-----A-------------G------T-T-----C--C--  2405
A_ROD         ----------C--C-----------------GA-C--T----------A-----T--A----------  2480
B_EHOA        ----------C-------------G-----G--T--A-----G-----T--------C---------C--  2537
SD_MM251      ----------C-----C---C-----------A-AG-------T-----A--------T--------C--  2540
STM_STM       ----G-----C---------C-------C--GA-A--T----------A--------T-----G-----  2495
VER_AGM3      ----------C-----------------C--G--A-------------C--------C---------C--  2516
GRI_AGM677    ----G-----C--C--C-----------------T--A----------A--------C---------C--  2513
SAB_SAB1C     ------------------CC-----------A-C--------G--T--T-----T--T----------  2492
SYK_SYK       ----G-----C-----T---GAG-----C-----A---TGT--------AA-CAC---T-------A---  2498

POL_CONSENSUS AGAagCagAagT?aT?cca??aGAaaCaGGaaaagaaACaGCata?TTtcTgtTaaAatTagcagg?AGa  2167
A_U455        ----------T--C---GC----------C-G--G--------C---A-AC-------------A---  2463
B_HXB2R       ----------T--T---GC---------GC-G----------T-----T-------------A---  2466
D_ELI         ----------T--T---GC---------GC-G----------T-----T-------------A---  2463
O_ANT70C      ----------G--A---GC----------C-------T--C--C--C---------C-G--T-CA---  2454
O_MVP5180     ---------G--G--A---GC----------C-G-----T--C--T--C----------------CA---  2454
CPZGAB        ----------T--C---GCT--G------C----------T--T--C----------------T---  2460
CPZANT        ------T-----A--GG--GAT--G-----G---AGT--------C--C-------------CA-C---  2475
A_ROD         ----------C--C---CAG---T------G-C--------CTC--C--A--G---C-G---A-T--G  2550
B_EHOA        ----------A--A--CCA----------G-C-G-----TCTC--C--------GC-G--CA-C---  2607
SD_MM251      ----------A--T---CA---G-------G-C-G------CTA-------------G-----C---  2610
STM_STM       ---G--------A--C---CA---------G-G-C-G-------TG--------G--G------A-C---  2565
VER_AGM3      ----------T--A--TAG---------G------------A-A---T-A-----GA--CT-A-T---  2586
GRI_AGM677    ---------G--G--C---AG------T--G--G--G------C-C------C-G---C-GTT--CA---  2583
SAB_SAB1C     ----------A--C---CAG-------------CT------C-C--C---C-G--G-----CA-C---  2562
SYK_SYK       ----A-CA-GA-CT-AAA-AGG-----G--GG-T----------TG-----CA-GC--A----CA-T---  2568

POL_CONSENSUS TGGCCagTaAaaca??TaCAcACaGAcAAtGG??c?AAtTTtac?agt?a?ga?gT?aaagc?gc?tgtT  2225
A_U455        ----------GTAA----------CAGC-----C--C--CGCT-CA--T-----A-TC----  2533
B_HXB2R       ----------ACAA----T--T--------CAGC-----C--CG--GCTACG--T-GG--C--C----  2536
D_ELI         ----------GTAG----T----------CAGC-----C--C---GCT-CA--T--G--C--C----  2533
O_ANT70C      -----T--T---GTAA----T---------C--GC-T--------A---ACAACTA-G--G--T--A----  2524
O_MVP5180     -----T--C---GTAA----T----------AC-T--------A---GCA-CCA-G-----T--A----  2524
CPZGAB        ----------ACTA-T--------T-----GC-A--------A---GCT-CA--C--G--T--C----  2530
CPZANT        ----------ACAA-------T--------AG-T-----C--A---GCA-CA--A-----G--A----  2545
A_ROD         ------A---C---CT-G--T-----T-----TG-C--C--C--TTCAC-G--G--G--GATG-TAGCA-  2620
B_EHOA        -----TA-C-C---CC-G----------C--TG-C--C--C--TTCAC-A--T--G---ATG--AGCC-  2677
SD_MM251      -----TA-T-C---TC---------T-----TG-T--C---G-CTCGC-A--A--A--GATG-TTGCA-  2680
STM_STM       -----T--C-C---CC-G-----G--T-----CG-T--C--C--CTCGC-G--G--A--GATG-TAGCC-  2635
VER_AGM3      -----CA---CC--AC-G--T----------AC-C--------GTC-C-A--A--AGC---AATG----  2656
GRI_AGM677    --------G-----TC-------T--T-----CC-A--C-----CTC-C-GA-T--GGC---G-TG--C-  2653
SAB_SAB1C     -----CA-C-C---GC-G--T-----T-----TA-C--C-----T--CC-GC-A--AGC---TATT--C-  2632
SYK_SYK       ------A-------GA-------T--T-----AC-A--C---GTC---G-CA-GT-C-----A--A----  2638

POL_CONSENSUS GGTGGgcagg?ATagAaCA??c?tttGGaaTaCC?TAcAAtCCacAaagtCAaGGagtagTaGAa?ccat  2289
A_U455        --------AAT--CC----GGAA-----G--C--C--------C----------------G---T----  2603
B_HXB2R       -------G--A--CA-G--GGAA--------T--C--------C--------------------T-T--  2606
D_ELI         ----------T--CA----GGAA--------T--C--------C--------------------T-T--  2603
O_ANT70C      -------CAAC---C----TGAG----------A--T-------------------------G----  2594
O_MVP5180     -----A----C---C----TGAG-----G-----A--T-------------------------G----  2594
CPZGAB        ---------AC--CA-G--GGAA----------C--T--------G----------G------T--T-  2600
CPZANT        -------TAAT--CC----GGAA----------A-------------------------G---T----  2615
A_ROD         -----AT---T--------AT-C------G----T----------G--C----------G-A--  2690
B_EHOA        -----AT---G--------AA-A--C---G-G--C--T------G-------G----------G-A--  2747
SD_MM251      ----------G-----G--CA-C-----GG----A----------G-----G--------G---G-A--  2750
STM_STM       -------G--A--------AA-C------G----T----------G--C----------G---G----  2705
VER_AGM3      ------G-AAA--------CA-CACA--TG----C--T--C--T---TCA-----CTCTA----GAGT--  2726
GRI_AGM677    ------GTAAT-----G--CA-CAC---------T--T--C-----GTCA--G--TAGT------AG---  2723
SAB_SAB1C     ------G-AAA-----G--CA-A-----GG----C-----C--C--------G----------T-T--  2702
SYK_SYK       -----TGT--C--------CA-TACA--T--T--A-----C--------C--G---A-------GT---A  2708
```

```
POL_CONSENSUS .gAAtaa?cA?tTaAAaaaaat?ATagg?caaaT?AGaGAtcAagcagaaca??ta?aaACagcaGTac?a   2350
A_U455        ---C--GG-A-----G-----C-----G--GG-A-----G-----T-----CC-TA-G----------A-   2673
B_HXB2R       ------AG-A-----G-----T-----A--GG-A--------G--T-----TC-TA-G----------A-   2676
D_ELI         ------AG-A-----G-----T-----A--GG-A----------T-----TC-TA-G----------A-   2673
O_ANT70C      ------GG-A------TC---T---CAG--GG-G--G--C-----------CT--AG----------A-   2664
O_MVP5180     ------AG-A------TCT--T---CAG--GG-G--G--C--------G--TT--A----------A-   2664
CPZGAB        A-----AG-GC----G-----A-----A--GG-T--G-------------TT--A----------A-   2670
CPZANT        ------A--A--G--GC----T-----A-----T-----C----------AT--A-G---------GT-   2685
A_ROD         ----C-C--TC-------CCAA---A-TAG---C-----A--G---A-TACAA--G-----AT-----T-   2760
B_EHOA        ---CC-T--TC-G-----TCAG----ACAG---T--------G----T-TCAA--G-G----TT--GTT-   2817
SD_MM251      ----C-C--CC-G-----TCAA----ATAG---C--G--A----A-TTCAG--G----CAT----TT-   2820
STM_STM       ----C-T--T-----G-CCCAG----ACAG---T----------A-CACAG--G----T-T-----TC   2775
VER_AGM3      ------A--G------G----A--T--GA----A------G-CTGTC--T-CAC-G----------TT   2796
GRI_AGM677    ---C-GA--GC-C--GG----C--CTCT-----A------G-TTGT--GAGAT-GG-G--------G-A-   2793
SAB_SAB1C     ------G--G------G----C-----A-----T------G-T------AGAT-GG----------CAT-   2772
SYK_SYK       A----GGT-T------G--GCT--CTCA-----T-----CG-C-TCAC---TT-GC----------GGC-   2778

POL_CONSENSUS ATGGCagt?cacatccacAATTTTAAaAgAAaagGgGGgaTtGGGGggtacact?caggagAaAGa?taa   2417
A_U455        --------ATT---T-------------------------------A----G-G----G------A---   2743
B_HXB2R       --------ATT-----------------------------------G-G----G------A--G       2746
D_ELI         --------ATT-----------------G-A----------A----G-G----G------A---       2743
O_ANT70C      --------ATTTG-T-------------------------------G----------GA---         2734
O_MVP5180     --------CTTTG-T-------------------------------G----G--G---C---         2734
CPZGAB        --------GTT---T-------------------------------G----G------A---         2740
CPZANT        --------G-----T-------------------------------AC-T---C-G---A--T         2755
A_ROD         ------A-T--TTG-ATG----------GG-----A--A---ATATG---C--TC-------T---     2830
B_EHOA        ------ACT---TG-ATG----------GG--A--A--A----ATATG--CC-T-C-------A--G     2887
SD_MM251      --------T--TTG-ATG----------GG--A--A--A----ATATG---C---C-------T---     2890
STM_STM       --------T---TG-ATG--------G-A--GG--A--AC-A----ATATG--CC---C-------T--G   2845
VER_AGM3      ------TGC-------------------G--A--A--A-------TA--AC-G-C---G---T---     2866
GRI_AGM677    -----TACG--T-----------------G--A--A--A-----TAT-T--AGT-C-------T-GG     2863
SAB_SAB1C     --------G-----T-------------------------------T--G---C-------C---     2842
SYK_SYK       ------ACCTT----TTA----------G--A--A--A------ATAT-AC----------GTAC-     2848

POL_CONSENSUS T?aAcATaaTa?caaCAgAa?tacaaacaa?a?aatta???cAaaaacaAattt?AAAAaTtcaaaAtTT   2478
A_U455        -AG--------G-------CA-------T-A-G-----...-------------C---------------   2810
B_HXB2R       -AG--------G-------CA-------T-A-G-----...-------------AC--------------   2813
D_ELI         -AG--------G-------CA-------T-A-G-----...-------------AT--------------   2810
O_ANT70C      -AG-----T--G--T--C--A---------C-G-----...-------------T----             2790
O_MVP5180     -AG-----C--G--T--C--A---------C-G--C--...-------------T------CA-C-----   2801
CPZGAB        -AG--------G-------CA---------GCG-----...-------------T----G-------A--   2807
CPZANT        -AG-----C--G-------CA----G----CTC-----...-----T-------T----------C-A--   2822
A_ROD         -C--T--G--CA-C------...---GAG-T-C----CCTC---GCCA---A--C----T-AA--G----   2897
B_EHOA        -C-----G---A-T------...---GA--T-C----CCTC----C-A---A--T----T-C--------   2954
SD_MM251      -T-----G--CA-T------...---GA--T-C----TCAA---TC-A---AC-C----T--A-------   2957
STM_STM       -T--T--G--CA-C------...---GA--T-C----TCAA---TC-A---A--C----T-CA-------   2912
VER_AGM3      -C--T--G--TA-T---C--...TT-GA-TT-C--CACCTA----CCA-----CA-------TT------   2933
GRI_AGM677    -T--T--GC--A-----C--...-T-GA-CT-A-TACTCTA-----C-----CCA-------TTG-----   2930
SAB_SAB1C     -C--T------CAT------C--G----C-A-AC-C-T...---C--A------C---------------   2909
SYK_SYK       -T--T--GC--TAT------...-T-CA-TT-C--CA-...A-T-C-AC-TCACC----T--TCG-----   2912

                          \/ 3'sj                                        5'sj \/
POL_CONSENSUS tcggGTctAtTaCaGagAagGcAgagAtcca?t?TGGAAaGGaCCaGca?a?cTacTgTGGAAAGGtGAa   2544
A_U455        ------T--------G--CA-------C--CA-T--------------A-A-----C------------   2880
B_HXB2R       ------T--------G--CA-----A-----C-T--------------A-G--C--C------------   2883
D_ELI         ------T----------CA----------A-T--------------A-G--C--C------------   2880
O_MVP5180     ------------------TA-------C--TA-T----------G---C-A--C--------------G   2871
CPZGAB        ------T----------CA-------C---A-T----------G---ACCT----------------   2877
CPZANT        ------TC-------G--TA-------C--TG-G--------------C-A--T---------------   2892
A_ROD         -----------T-----------------AGT-G----------T-GGG-A--------------A---   2967
B_EHOA        C-----------------------------A-C-C-----G-----T-GTG-T---T----------G---   3024
SD_MM251      ----------------------------A-C-G-----G-----C-GTG-G---T----------G---   3027
STM_STM       ---------------G------------A-C-A-----G-----C-GTG-G---T----------G---   2982
VER_AGM3      -A-A-----C----------G--------TG-C-------------G-C-GT--A-T--------G---   3003
GRI_AGM677    -AA------C----------T---------G-G-------------GCGA--CA-C--------A---   3000
SAB_SAB1C     ---------------G----------C--TG-G--------T--G--TA-G--CA-C--------A---   2979
SYK_SYK       -A----T--C---C-CC-G--G-AGA-CGAG...-----G--------TCGT-----T----------   2979
```

```
POL_CONSENSUS GG?GCaGTagT?aTacAa???Ga?gagagaGAcaT?aaagtagtaCCaAGaAGaAAaGcaAAgAT?AT?A   2605
A_U455        --G--------A------...--CA-T--T--T--A--G-------------------------C--T-   2947
B_HXB2R       --G--------A------...--TA-T--T-----A--------G-------------------C--T-   2950
D_ELI         --G--------A------...--CA----T-----A--G--------------------T------T--T-   2947
O_MVP5180     --G--------C------...--TA-AG-------T-----G--------------G-----A--A--C-   2938
CPZGAB        --G--------G--C--G...--TC-AG-G--AC-A--G--------------G-----------C--T-   2944
CPZANT        --G--------A---                                                          2907
A_ROD         --A-----CC-AG-CA-G...-TA-GA-C------A---A--A----------G-----C-----C--C-   3034
B_EHOA        --A-----CA-C---A-G...-TA-G--C---A--C------A----C----------------C--A-   3091
SD_MM251      --A-----CA-CT--A-G...-TA-G--C------T--G--------C--------G--T--A--T--C-   3094
STM_STM       --A-----CA-CG--A-G...-TA-G--C------T--------------G-----G--T--A--C--C-   3049
VER_AGM3      --T-----G--C--CA--...-GA-GTGTG--AT-A----A-TAC--------G--------A--T--A-   3070
GRI_AGM677    --C--G--G--A--TA--...--G-G-GA------C--G-----C--C--G-----G--T-----T--C-   3067
SAB_SAB1C     --A--------T------...--AC--G-T--AT-G---AC-A-C--------------C--A--T--A-   3046
SYK_SYK       --T--------GG-G---ACA--G---G-------TTTT-C---------------G-----A--A--T-   3049

          vif start ->
POL_CONSENSUS gagAttATGgA?aa?gaatgg?aggtgatg?t?gt?tg??????g??g??a??aga?aggct??????gA   2650
A_U455        -G---------A--CAG----C--------A-T--A--.........-CAGGT---C---A-......--   3002
B_HXB2R       -G---------A--CAG----C--------A-T--G--.........-CA-GT---C---A-......--   3005
D_ELI         -G---------A--CAG----C--------A-T--G--.........-CA-GT---C---A-......--   3002
O_MVP5180     ----------A--CAG----C----AC--A-A--A--.........-CA-AT---C--A-A......--   2993
CPZGAB        ----------A--CAG----C--------A-T--G--.........-CA-GT---C--AA-......--   2999
A_ROD         ----C------GG-A--CAA-AGATG---AG-G--TCCCACCTG-AG-GTGCC--GG---A-GGA...--   3101
B_EHOA        -GA-C------GG-G---AA-A-TTG---TGCA--GCCGACGTG-AG-AT-CC-TGC-----AGA...--   3158
SD_MM251      A----------GG-G---AA-AG-TG---AGCA--TCCCACATG-AG-AT-CCG--G-----AGA...--   3161
STM_STM       A----------GG-G---AA-AG-TG---AG-G--TCCCACCTG-AG-AT-CC--GG-----GGA...--   3116
VER_AGM3      AG-------A-CC-A---AAAG-ATG-G--A-GAGAGTAACTTG-AA-GTGCCG--GGA---.......--   3134
GRI_AGM677    A----------G-GA---AAAC-ATG---AG-GAGGGTAGTATG-AG-GTGTC---G----AAATAAGC-   3137
SAB_SAB1C     A----------A--GC-T---ATA--C-G-CCCC-C--...GAA-GTAACGGGC-GACA--AGGA...--   3110
SYK_SYK       C---CC-----G--A------ATA---G-TCCCACG--..............GAGA-T-ACCCC...A-   3101

POL_CONSENSUS g???ga?gag??tA?tGA????C??G?T?A???ACCA                                  2669
A_U455        A...--T                                                                 3006
B_HXB2R       -...--T                                                                 3009
D_ELI         -...--T                                                                 3006
O_MVP5180     A...AGT--AAGC-TG--ACAG-CT-G-G-AAT----                                   3027
CPZGAB        -...--T                                                                 3003
A_ROD         AATG-CA                                                                 3108
B_EHOA        -GTG-CAC--TC--A-                                                        3174
SD_MM251      -GTG-CA                                                                 3168
STM_STM       -GTG-CA                                                                 3123
VER_AGM3      T...A-C                                                                 3138
GRI_AGM677    -ATG--G-G-GA--G---CTTA-AA-A-C-GGA-                                      3171
SAB_SAB1C     -GTG--C                                                                 3117
SYK_SYK       -ACA--T                                                                 3108
```

HIV1 POL CONSENSUS

```
        gag ->
CONSENSUS-A   TTTTTTAGGGAA...AATTTGGCCTTCCA?CAAGGGGAGGCCAGG?AATTT.....................     46
CONSENSUS-B   ------------...g--C---------ca------a--------g-----????????????????????     48
CONSENSUS-D   ------------...g------------CA------A-----G--G--c--.....................     48
CONSENSUS-O   ---------?--...?TAC------C-GGGGGGCAC---------C----A.....................     46
CONSENSUS-U   ------------...-------------CA------A--------G-----.....................     48
CONSENSUS-CPZ ------------???????????-?-???G??G??-?--?--------C--.....................     29

CONSENSUS-A   .................TCCTCAGAG......CAGACC?GA?CCAACAGC?CCACCAGCAGA?A?CTTTG     88
CONSENSUS-B   ??????????????????--t------.....------A--g--------c--------a---G-G---Ca     95
CONSENSUS-D   .....................--T------.....------A--G--------C-----------G-G---C-     95
CONSENSUS-O   .................-GTG----?......AC?-G?-??C---T----C-----?AT..........     75
CONSENSUS-U   .................C-T------.....------A--G--------C-----------G-G---C-     95
CONSENSUS-CPZ .................-GTG----A?????????-??---------?--------?AT---G--?-A-?     63

CONSENSUS-A   GGATGGGGGAA?AGAT??C???...CTC?CC.........?????A?CAGGA?C?GA??GA......CA?    122
CONSENSUS-B   --T-t------G---caa-AaCTCC---T-A?????????---GA-G-----g-c--tA--.....--A    147
CONSENSUS-D   --T-T------G----...AACCCC---T-A...........gA-A-----G-A--AA--.....--A    144
CONSENSUS-O   ....--A---GG?AG-GAA..............GG?ACA-GAGA-T---??--A--A--GGG??C?G?A    115
CONSENSUS-U   --T-T-----GG----AAA...ACC---T-A...........GA-A-----G-A--AA--.....--A    144
CONSENSUS-CPZ ?-?A??A-?-?--???GA?---??????-??.....---?-?????-?-??--?????????--     72

                                          \/ protease                          <- gag cds
CONSENSUS-A   GG????.....A?A??CTCCTTTA??TTCCCTCAAATCACTCTTTGGCAACGACCC?T?GT?ACAGTAA    174
CONSENSUS-B   --a.........-cTGTA-------gC--------g----------------------c-c--C---a---    208
CONSENSUS-D   --A.........-CTGTA-------aC------------------------------C-t--c---A---    205
CONSENSUS-O   ?-A.........GCT-TA?--?--TGCC-----------C---------G-CA----AA-A-------C--    171
CONSENSUS-U   --A.........-TTGTA-------GC----------------------------AG-T--C-------    205
CONSENSUS-CPZ ??G---???????CTA-?---??C?--?-----?-----?----------??----?---A--??---??    110

CONSENSUS-A   ?AATAG??GGACAGCT?ATAGAAGC?CTATTAGA?ACAGGAGCAGATGATACAGT?TTAGAAGACATAAA    237
CONSENSUS-B   ag----gG--G--A--a-Ag-----T--------T--------------------A--------a--G--    278
CONSENSUS-D   A-----GG--------A-AG-----T--------T--------------------A--------A--g--    275
CONSENSUS-O   -?G-T-GG--?--?--ATGT--?-?TT-?C-G--T-----?--------?-----A?--A-?A-----C-    230
CONSENSUS-U   G-G---GA--------A-A------T--------C--------------------A--------A-----    275
CONSENSUS-CPZ -??--?A---?--A??ATGT?-?--T?-??----T--------T-----?-----A?----G?GA--?C-    163

CONSENSUS-A   TTT?CCAGGAAAATGGAAACCAAAAAT?ATAGGGGGAATTGGAGGTTTTATCAAAGTAAGACA?TATGAT    304
CONSENSUS-B   ---G-------g---------------G-----------------------------------g------    348
CONSENSUS-D   ---G-----------------------G-----------------------------------G------    345
CONSENSUS-O   A---GA-----?-----?---------G--------T--A-----?-----A-------A-G-----?-?    293
CONSENSUS-U   ---G-----------------------G-----------------------------------G------    345
CONSENSUS-CPZ ?----A----???-----?-------?---?--?--------?--------????--??A-----?-?-?    212

CONSENSUS-A   CA?ATACTTATAGAAATTTGTGGAAAAAAG?C?ATAGGTACAGTATT?GTAGGACCTACACCTGTCAACA    370
CONSENSUS-B   --g-----c---------c------C-T--AG-T--------------A----------------------    418
CONSENSUS-D   --A-----C---------C------c-t--AG-T--------------A----------------------    415
CONSENSUS-O   A-TG-GACAG-----?-A?AA----?GG-?GTACAG--A--------G--G--------T-----T--T-    359
CONSENSUS-U   --A-------------------------G-T---------A----G----------------------    415
CONSENSUS-CPZ A--G---?C---?-?--?G?A-??-G?-?AGTA?---??------?--?-?-----A-??--?--A--?-    261

                                                  protease \/ p66, p51
CONSENSUS-A   TAATTGGA?G?AATATGTTGACTCAGATTGGTTGTACTTTAAATTTTCCAATTAGTCCTATTGAAACTGT    438
CONSENSUS-B   --------A-A---c-------------------c--------------c----------------------    488
CONSENSUS-D   --------A-A---T-------C--------C--C-------------------------------------    485
CONSENSUS-O   -T?----?A-A--?--A-----AGGAT-A--?-----AC----?--C--T--A--?--C--A-CCC-A--    423
CONSENSUS-U   --------C-A-------------------------------------------------------G-----    485
CONSENSUS-CPZ -?--?--?A-A---?-T--????-??T-?--?------?--???--?---------??A?-------?--    310

CONSENSUS-A   ?CCAGTAAAATTAAA?CCAG?AATGGATGGCCCAAAGGTTAAACAATGGCCATTGACAGAAGA?AAAATA    504
CONSENSUS-B   A--------------G----G--------------A-------------------------a------    558
CONSENSUS-D   A--------------G----G--------------A-------------------------A------    555
CONSENSUS-O   G---------?C----A----G---------A-----A--A----------CC-AT-TA?----------    490
CONSENSUS-U   A--------------G----GG-------------G-------------------------A------    555
CONSENSUS-CPZ ---?--?-----?--A??--G---------?----??--A--?---------??-?T--??---------?    363

CONSENSUS-A   AAAGCATTAACAGA?ATTTGTA??GAGATGGAAAAGGAAGGAAAAATTTCAAAAATTGG?CCTGAAAATC    570
CONSENSUS-B   ---------gt---a-------Ca--a---------------g----------------g----------    628
CONSENSUS-D   --------------A-------CA--a-------------------------G------G----------    625
CONSENSUS-O   G----??-?--?-CA--A---CA---A------C-?----------?----G---A--A----------    553
CONSENSUS-U   --------------A-------AA--T----------------------T---------G----------    625
CONSENSUS-CPZ ?-?--C?---?---A------?A-?-??-?---???---???--?--?--?-?---A-----?--?--?-    409

CONSENSUS-A   CATACAATACTCCAGTATTTGC?ATAAAGAAAAAGGA?AG?ACTAAATGGAG?AAATTAGTAGATTTCAG    636
CONSENSUS-B   ----------------------c----------a--c--T----------a----------------    698
CONSENSUS-D   --------------A-------C----------A--C--T--C--G-----A----------------    695
CONSENSUS-O   -T--T-----A--TA-?-----T-----A--G--A--T?-------?-----A-----G-----?--?--    617
CONSENSUS-U   ----------------------C----------A--C--C----------A--------GA-------    695
CONSENSUS-CPZ ----------?--?A-T-----A--?--A--G--A--C-?T?-?--------A--??----?--?-----    469
```

```
CONSENSUS-A   AGAACTCAATAA?AGAACTCA?GA?TTCTGGGA?GT?CA?TTAGGAAT?CC?CAT?C?GCGGG??TAAAA   694
CONSENSUS-B   ------t-----g--------a--c---------A--t--a---------A--A---C-c--a--gt-----   768
CONSENSUS-D   ------T-----G--------A--T---------a--T--A---------A--G---C-t--A--GC-g---   765
CONSENSUS-O   ?---T-A-----------A--A-----------G------?----T-----A---C---G---TT-?--G   678
CONSENSUS-U   ---G--T-----A--------A--T--T-----A--T--A--------A--A---C-T--T--GT-G---   765
CONSENSUS-CPZ ----?-A-----------A--A-----T------?----------?-----T--?C-A--?---T----?   526

CONSENSUS-A   AAGAAAAAATCAGTAACAGT?CTAGATGTGGGGGA?GCATATTTTTCAGTTCC?TTAGAT?AA??CTTTA   758
CONSENSUS-B   --------------------a--g--------t--t-----------------c---------ga---c-   837
CONSENSUS-D   --------------------A--G--------T--T-----------------C------g--GAt----   835
CONSENSUS-O   C-A-?GC----T--T-----CT-------A--A--T--T-----C---TG?--------?CC-GA?----   743
CONSENSUS-U   --------------C-----AT-G----------T--------------C--T------G--GAT--C-   835
CONSENSUS-CPZ ?-?--?--------G-----AT-?-----?--A--T--C--?--C--????--C?-?-----?GA?--?-   580

CONSENSUS-A   GAAAGTATACTGC?TT?AC?ATACCTAGT?TAAA?AATGAGACACCAGG??T?AG?TA?CAGTACAATGT   818
CONSENSUS-B   -g----------A--T--C---------a----C--------------ga-T--a--T----------   907
CONSENSUS-D   -G--A-----c--a--T--C---------A----c--------------GA-T--A--T----------   905
CONSENSUS-O   ----A--?--------C--T--T------G-G-----------C-----A--A--A--C----------   809
CONSENSUS-U   -G----------A--C--T-----C---A-T--T-------------GA-T--A--T----------   905
CONSENSUS-CPZ -?--A-----?--------A--T--?--?--?-----------C-----A-----A--T???--T-----   636

CONSENSUS-A   GCTTCC?CAGGGATGGAAAGG?TCACC??CAATATT?CAG???AGCATGACAAAAATCTTAGAGCCCTTT   880
CONSENSUS-B   ------a--------------a-----AG-------c--aaGt----------------------t---   977
CONSENSUS-D   ------A--------------A-----GG-------C--AAGT--------------------------   975
CONSENSUS-O   C--C--G--A--?--------T-----AG-?--------?AGTTCA--------?--TC----T--A--?   873
CONSENSUS-U   ---A--A-------------A-----AG-------C---AGT----------------------A------   975
CONSENSUS-CPZ T?-?--A--A--?----------?--A--T--?-----A-----?-----?--?--??---?A--?--?   687

CONSENSUS-A   AGA?CA?AA?ATCCAGA?ATAGT?ATCTA?CAATACATGGATGA?TT?TATGT?GGATCTGATTTAGAAA   941
CONSENSUS-B   ---AA-C--A-------c-----t-----t--------------t--G-----A--------c-------  1047
CONSENSUS-D   ---AA-C--A-------A-----T-----T--------------T--G-----A--------C-------  1045
CONSENSUS-O   --?A?--?CA-C-----A?---AA--T-?T--G-----?-----C?-A-----A-----A------CC??  933
CONSENSUS-U   ---A--A--A-------A-----C--A--C--------------T--G-----A--G-------------  1045
CONSENSUS-CPZ --?GA?A-G-----?-?-?-???---T-----?----------C----------?------?-?----   738

CONSENSUS-A   TAGGGCA?CATAGAGCAAAAATAGAGGA?TT?AGAG??CATCTA?TGA??TGGGGATT?A??AC?CCAGA   999
CONSENSUS-B   -------G------a----------ac-G---caa-----gt---gg--------t-cc--a-----  1117
CONSENSUS-D   -------G------A----------A--A----AA------T---GG--------T-CC--A-----  1115
CONSENSUS-O   -G?CAG-A------AA?-GG?-?--ATTGC-T--?-AA--??--TATCAG----------CT--C--T--  995
CONSENSUS-U   ----A--A------A----------AC-A----AA------T---AA--------T-CC--A-----  1115
CONSENSUS-CPZ -T??T??A------?A--?G?-?--A?--C-----CAA---T-?C???-------??--?-------?--  783

CONSENSUS-A   CAAAAAGCATCAGAAAGAACCTCCATTTCTTTGGATGGG?TATGAACT?CATCCTGA?AAATGGACAGT?  1065
CONSENSUS-B   ------a----------c----------T--------c--------t----------a  1187
CONSENSUS-D   T-----A----------c----------t--------C--------T----------A  1185
CONSENSUS-O   ?----------?--G--------C---?-A----------G--C--?--A--C--?--------A  1059
CONSENSUS-U   ----------C----------G--------C--C-----C----------G  1185
CONSENSUS-CPZ ------?-----?--?--?--?--?---??A--------A-----G?----?--A--C--------?---  840

CONSENSUS-A   CAGCCT?TA?A?CT?CCAGAAAAAGACAGCTGGACTGTCAATGATATACAGAAATTAGT?GGAAAACTAA  1130
CONSENSUS-B   ------A--gtg--g----------c--------g-----G------T-g-  1257
CONSENSUS-D   ---t--A--A-a--G--------g--g----------g-----G--G---T---  1255
CONSENSUS-O   ---?-CA-CC-AT-G--T?-C--?--?GTG-----A--A----------A---?--?-A-----??---  1121
CONSENSUS-U   ------A--C-A--G-----C--G--A----------G--G----------  1255
CONSENSUS-CPZ ---???A-?---?-A-----???---?G??-----?--?-----?--?--?---C-??-A-----?T---  889

CONSENSUS-A   ATTGGGCAAGTCAGATTTATGCAGG?ATTAAA?TAAAGCAACTGTGTA?ACT?CTCAGGGGAGCCAAAGC  1196
CONSENSUS-B   ----------G------G--------t-A----a---c--T------a-------  1327
CONSENSUS-D   -C---------C---------c----A------G---g----T-a----A---C--T------A-------  1325
CONSENSUS-O   ----------A--C---CA---A----G-G-?-?AG--T----?-AGT-AA-?--A--?A----?T-  1185
CONSENSUS-U   ----------C----A------G--------T-A----A---C--T--------A-----  1325
CONSENSUS-CPZ ----------?--?C----A--?---A??-----?T-?-----?---A-?--A---???--GA?  944

CONSENSUS-A   ACTAACAGATATAGTA?C?CTGACTGA?GAAGCAGAATTAGAATTGGCAGAGAACAGGGA?ATT......  1256
CONSENSUS-B   ---------Ag--a--c-a--a--A--A--------gc-----c-------A--------g---.....  1391
CONSENSUS-D   ---------AG--a--C-A--A--A--A---------------C-------A--------A---.....  1389
CONSENSUS-O   -T-?-----?G-----C-TT-A-G-A?A--G-----?C-?-----A-A?--A-----A--A-?G.....  1242
CONSENSUS-U   ---------C------C-AT-A----CA--G----------------------------A---.....  1389
CONSENSUS-CPZ ?--?------??---?C?-?-?--???----------------A-?---A--T-???----???????  989

CONSENSUS-A   ...CTAAAAGA?CCTGT?CATGG?GTATATTATGACCCA?CAAAAGAC?TAGTAGCAGAA?TACAGAAAC  1317
CONSENSUS-B   ...--------a--a--a-----A--g----------t--------T--a--------a-------g-  1458
CONSENSUS-D   ...--------A--A--A-----A--G----------T--------T--A--------a---------  1456
CONSENSUS-O   ...?-----?----A----------?--?C-?--TGA?--?--?--?TGG-TTA?TA-T-----G-  1295
CONSENSUS-U   ...--------A--A--G-----G----------T--------T--A--------A-------G-  1456
CONSENSUS-CPZ ????--?????-??A?-A?-?---???--C---?-?--?G????-???C-?A?----???A---------  1025
```

HIV1 POL CONSENSUS

```
CONSENSUS-A   AAGGGCAAGACCAATGGACATATCAAATTTATCAAGAGCCATTTAAAAAT?T?AAAACAGGAAAATATGC   1385
CONSENSUS-B   -g-------g-----------------------------------------c-g----------------   1528
CONSENSUS-D   -------C-G-----------C--------------A-----------C-g----------G-----   1526
CONSENSUS-O   -?--AG??-?G--------T--C--G?-A-----G--?GA-CA---G--CC----------?------?-   1356
CONSENSUS-U   -G-------GT-----------------A--C-------A--A-------C-G--------G--G-----   1526
CONSENSUS-CPZ -?--?????G?--?-----?-----?--A-?---??-???-???--????T----??----?--------   1068

CONSENSUS-A   AA?AAA?AGGTCTGC?CACACTAATGATGTAAA?CAATTAACAGA?GT?GTGCAAAAAGT??C?A??GAA   1444
CONSENSUS-B   --g--tg---GG---C-----------------A----------G-CA---------a-ag-C-ca---   1598
CONSENSUS-D   --G--tG---GG---C-----------------a------g----G-cA---------A-At-C-CA---   1596
CONSENSUS-O   T-GGC-A-A-G-CT-C-----A------A---GA-----?G----A--A?-C--G--G--GT-TCAA---   1424
CONSENSUS-U   --G--TA-A------C----------------A----------A-CA--------GA-AG-CCAA---   1596
CONSENSUS-CPZ ?-G?C?--????A?-------?-----??-?-GG---?-?G-?-?A-?A--?------A--G?-?-----   1112

CONSENSUS-A   AGCATAGTAATATGGGGAAAGA?CCCTAAATTTAGACTACCCATACAAA??GA?ACATGGGAA?CATGGT   1509
CONSENSUS-B   ----------------------ct----------a----------Aa--a---------a------   1668
CONSENSUS-D   --------G----------CT----------------------AG--A---------A------   1666
CONSENSUS-O   ?CT---?-T--?-----?--ATT?--------?-?G--G--AG-?ACT-GA--A--T------A-?----   1485
CONSENSUS-U   ----------------A-CT----------------------AA--A--------GG------   1666
CONSENSUS-CPZ --?-----?--?------?-??-A--?------???T----??--???-----A-?T-----?G-?---   1161

CONSENSUS-A   GGATGGAGTATTGGCAGGCTACCTGGATTCCTGAATGGGAGTTTGTCAATACCCCTCC?CTAGTAAAATT   1578
                                                                        /<- CRS
CONSENSUS-B   ---ca----------A--c----------g----------cT----g-----   1738
CONSENSUS-D   --g-A----------A--C--t----------G-----A----------TT----------   1736
CONSENSUS-O   --GC?--?--------?--C----------?--------A-------GC--A--C--AT-GA-C-----   1551
CONSENSUS-U   ---CA--A--------A--C--------C----------T-----C---------C-   1736
CONSENSUS-CPZ --?CA--?--?--------?--------?--?--------?---?-?-----?--?---T--?-?-?-?-   1215

CONSENSUS-A   ATGGTACCAGTTAGA?AAAGACCCCATAG?AGGAGCAGA?AC?TTCTATGTAGATGG?GCAGCTAATAG?   1642
                                  - - - CRS - - -
CONSENSUS-B   ---------------g-----a-------t---------A--t----------g----------g   1808
CONSENSUS-D   ---------------G--G--A------AT---------A--T----------G----------A   1806
CONSENSUS-O   -------??-?----A-??--A--T--T?T?--G------A--C-?T----------A----------   1612
CONSENSUS-U   ---------------A-C---A-------T---------A--T----------G----------A   1806
CONSENSUS-CPZ ?-----?A?T?--??G?C---?--T--?CC-????-?-------?T-----?----------?------   1264

CONSENSUS-A   GA?ACTAAG?TAGGAAAAGCAGGGTATGTCACTGACAGAGGAAGACAAAAGGTTGTTTCCCTAACTGAGA   1710
                                  - - - CRS - - -
CONSENSUS-B   --g-----At----------A-----t----------a-----c----------C-   1878
CONSENSUS-D   --G-----AT----------A-----T----------G--A-----CC-TT-?-----c-   1875
CONSENSUS-O   ?----A--AC-------G--?--A-----T--A--ACA---?-A---G--?A-AA-AAA?T--GA?----   1675
CONSENSUS-U   --A-----AAAG----------A-----T----------C---T-------A-   1876
CONSENSUS-CPZ ?--?-??-A-??--G--?-----?-----?--A----?---?-?-???-???-?A????-?--?AAA-?-   1306

CONSENSUS-A   CAACAAATCAAAAGACTGAATTACATGCAAT?CAT?TAGCCTT?CAGGATTCAGGATCAGAAGTAAATAT   1777
                                  - - - CRS - - -
CONSENSUS-B   ----------G--------g-----A-----T---c----t--G--------g----t---------c--   1948
CONSENSUS-D   -g--------G--------G-----A-----TA--c----t--g--------G----T---------C--   1945
CONSENSUS-O   -?--C--------?G--------ATG--??-ATTA----------CAA?GA???------?--   1734
CONSENSUS-U   ----------G----------A-----C--CT----T--A----------C--   1946
CONSENSUS-CPZ -C--?-----?C-AG-A------?-G--??---T---G--??--?--------??????C??--?--?--   1355

CONSENSUS-A   AGTAACAGACTCACAGTATGCATTAGG?ATCATTCAGGCACAACCAGACAGGAGTGAATCAGA??TAGTC   1844
                                  CRS ->/
CONSENSUS-B   ----------A----------a--------a----------T-aa----------gt-----   2018
CONSENSUS-D   ---------T-----A----------A--------A----------T-A----------GT-----   2015
CONSENSUS-O   ---------?-----A----?---G--C?----?TCCT-C-----?ACACA------?--CCCTA----T   1798
CONSENSUS-U   ----------G--T-----A----------T-AA----------GA-T--T   2016
CONSENSUS-CPZ ---?--?--?-----A----??--?---?-??-?--???????-----TCA?---------??------?   1401

CONSENSUS-A   AATCAAATAATAGAGAAGCTAATA?AAAAGGA?AAAGTCTACCTGTCATGGGTACCAGC?CACAAAGGGA   1911
CONSENSUS-B   -g----------c--t-----A-------a--g---------g----------A--------a-   2088
CONSENSUS-D   -g----------C--t-----A-------A--G--T------G----------A--------A-   2085
CONSENSUS-O   C-G--?---------G-A----C-A-------AC??--G--T--TA-------T--T-----?-----C-   1863
CONSENSUS-U   ----------C-AT-----C-G-----C--G----------A----------   2086
CONSENSUS-CPZ ?-??-?-----??-??-?T----?A---?A--A?-?A-?-----C--C----------A--?-----?-   1455

CONSENSUS-A   TTGGAGGAAATGA?CAAGTAGATAAATTAGTCAGT??TGGAATCAGGAAGGT?CTATTTTTAGATGG?AT   1976
CONSENSUS-B   ----------a----------gC----------A--a----------a--   2158
CONSENSUS-D   ----------A----------CAg----------A--A--------G-----A--   2155
CONSENSUS-O   -A----------AA--A----------A--CAA?-A?--T--A-GA--C--G--CC-?--A--A--   1930
CONSENSUS-U   ----------A----------CAG---------A-----A----------G--   2156
CONSENSUS-CPZ -A----------?--?--?----------A--?-----??-A-----?--?C-?--?-----   1510

CONSENSUS-A   AGATAA?GC?CAAGA?GA?CATGAAA?ATATCAC?GCAA?TGGA??GCAATGGCTAGTGATTTTAATCTG   2037
CONSENSUS-B   ------G--c-----a--a-----g-A-------A-T--t----gA----------c---   2228
CONSENSUS-D   ------g--T-----A--A-----G-A-------AA---T----GA----------C--A   2225
CONSENSUS-O   ---?C-G--A-----A--T-------A-------TA-T--T-----A---?-A--------?---GGA?-A   1995
CONSENSUS-U   ------G--T-----A--A-------A-------A----T----GA----------C--------A   2226
CONSENSUS-CPZ ---?-?----------?--?-----TA-?--T-----A?-??-?--???---?-??---?-?   1557
```

```
CONSENSUS-A   CCACCT?T?GTAGC?AA?GAAATAGT?GCCAGCTGT?ATAAATGTCA??T?AAAGGGGAAGCCATGCATG  2098
CONSENSUS-B   ------g-A-----A--A--------A---------g----------GC-A-----a-------------  2298
CONSENSUS-D   ------G-G-----a--A--------A--T------G----------GC-A-----A-------------  2295
CONSENSUS-O   -----A-----?--C--G-----CA-T--T--?---CC------?--TA-A-----------A-??----  2058
CONSENSUS-U   ------A-A-----G--G--------A---------G----------AC-A-------------------  2296
CONSENSUS-CPZ --?--?A----?-----A-----??----?CA?---G----?--?---G-A--?--A-------?-----  1612

CONSENSUS-A   GACAAGTAGACTGT?GTCCAGG?ATATGGCAATTAGATTG?ACACAT?TAGAAGG?AAA?T?ATC?TAGT  2160
CONSENSUS-B   --------------A-------A---------C-------t------t-------A---g-T---c-G--  2368
CONSENSUS-D   --------------A-------A-----------------T------t-------A---G-T---C-G--  2365
CONSENSUS-O   -T--?-------?CA-?----A-?--------A-?--------------?-----C---A-C--AA----  2119
CONSENSUS-U   --------------A-------G-----------------C------C-------A---A-A---A----  2366
CONSENSUS-CPZ -?--?--------?A-?--------?------G----------C--?------------G-----A----  1671

CONSENSUS-A   AGCAGTCCATGTAGCCAGTGGCTA?ATAGAAGCAGAAGTTATCCCAGCAGAAACAGGACAGGAGACAGCA  2229
CONSENSUS-B   ------t--------------a--t----------------T--------g-----g-----a------  2438
CONSENSUS-D   ------T----------------T----------------T--------------G-----A------  2435
CONSENSUS-O   T--T--------?--A----??-T-----------?--G--A----------------?--A--T--C  2183
CONSENSUS-U   --------------------A--T-------------------------------------------  2436
CONSENSUS-CPZ ?-----?--?---?-?-------?---------?-----?--??---?T--G-----??-A???-----?  1723

CONSENSUS-A   TACTTTATA?TAAAATTAGCAGGAAGATGGCCAGT?AAAGTAATACACACAGACAATGGCAGCAATTTCA  2297
CONSENSUS-B   ------c-ct-------------------------A---ac------t-------------------  2508
CONSENSUS-D   --t---c-tT-----------------------A------G----T-------------------  2505
CONSENSUS-O   --?--CC-GT-----?-?--?-C---------T-------------T--------?--?CCT-----T-  2246
CONSENSUS-U   ---------C-------------------------A------G-------------------------  2506
CONSENSUS-CPZ --?--CC-GT----------??-?----------A---AC?--?-----?--?-----??C?-----?-  1781

CONSENSUS-A   CCAG?GCTGCAGT?AA?GC?G??TGTTGGTGGGCAAAT?TC??ACA?GAATTTGG?AT?CCCTACAATCC  2355
CONSENSUS-B   ----Ta--A-g--T--g--C-CC----------GGga--AAG--G--------c--T----------  2578
CONSENSUS-D   ----T--------T--G--c-CC----------GG-A--AA---G--------A--T----------  2575
CONSENSUS-O   -A--T?-A?-?A-G-----T-CA---------?-???CA-ACA---T--G--------A--A--T-----  2307
CONSENSUS-U   ----T--------T--A--A-CC--------------A--AA---G--------A--T--------C--  2576
CONSENSUS-CPZ -A--T--?-------------C----------??-?A---A?--G--------A--A--?--?-----  1839

CONSENSUS-A   CCAAAG?CAAGGAGTAGTGGAATC?ATGAA?AA?GAATTAAAGAAAAT?ATAGG?CAGGT?AGAGA?CAA  2417
CONSENSUS-B   ------T----------A-----T-----T--a-------------T-----a-----A-----T--g  2648
CONSENSUS-D   ------T----------A-----T-----T--A-------------T-----A-----A-----T---  2645
CONSENSUS-O   A-----T----------A---G-C-----T----------ATC?--T---CAG-----G--G--C---  2375
CONSENSUS-U   ------T----------T-----T--G-------------C-----G-----A-----G---  2646
CONSENSUS-CPZ A--?--T--------?--?-----C?-?--T--A?-??-?---?----------A--??-T--?------  1894

CONSENSUS-A   GCTGAACACCTTAAGACAGCAGTACA?ATGGCAGTATTCATTCACAATTTTAAAAGAAAAGGGGGGATTG  2486
CONSENSUS-B   --------t----------------a-------------C-----------------------------  2718
CONSENSUS-D   --------T----------------A-------------C-----------------------------  2715
CONSENSUS-O   --A--?--?T-A-?A----------A--------?--TG-----------------------------  2441
CONSENSUS-U   ------------------------A--------G-------------------------------------  2716
CONSENSUS-CPZ --A-----?T-A--?---------??A--------G??-----------------------------  1958

CONSENSUS-A   GGGG?TACAGTGCAGGGGAAAGAATAATAGA?ATAATAGCA?CAGA?ATACAAACTAAAGAA?TACAAAA  2551
CONSENSUS-B   ----g-------------------g----c---------a----c---------------t-------  2788
CONSENSUS-D   ----A-------------------------C---------A----C---------------T-------  2785
CONSENSUS-O   ----G----C------?--?--??-------C---?-----T--C-A--------A-C------------  2505
CONSENSUS-U   ----G-------------------------C--G------A----C---------------T-------  2786
CONSENSUS-CPZ ----G----C??-?--??-?------?----C---?-----A----C-----?--A-???--T-------  2016

                                                    \/ 3' sj
CONSENSUS-A   ACA?ATT??AAAAATTCAAAATTTTCGGGTTTATTACAGGGACAGCAGAGACCCCATTTGGAAAGGACCA  2618
CONSENSUS-B   ---A---AC-----------------------------------------------t--ac----------  2858
CONSENSUS-D   ---A--cAt-----------------------------------------T--A---------------  2855
CONSENSUS-O   ---A---TT----?-??-C--?--------C--------A--?-----------T--?-----------G  2569
CONSENSUS-U   ---A---AC-----------------------------------A---------A---------------  2856
CONSENSUS-CPZ ?--A---TT----?-----?-A---------?-------?--?-----------??-?-----------?  2076

                    5' sj \/
CONSENSUS-A   GCAAAACTACTCTGGAAAGGTGAAGGGGCAGTAGTAATACA?GACAATAGTGATATAAAGGTAGTACCAA  2687
CONSENSUS-B   -----G--t-------------------------------a--T--------c-----A-----G----  2928
CONSENSUS-D   -----G--c-------------------------------A----------C---------------  2925
CONSENSUS-O   ---C-?--C--G----------G--?--------C-----A--T--?G-A--C--T--?--?-------  2634
CONSENSUS-U   ---------------------------------------G-------------------------------  2926
CONSENSUS-CPZ ---????-?--G----------------------?--?--G--TC-AG-G--AC---------------  2139

                              vif cds ->
CONSENSUS-A   GAAGAAAAGCAAA?ATCATTAGGGATTATGGAAAACAGATGGCAGGTGATGATTGT?TGGCAGGTAGACA  2755
CONSENSUS-B   ----------g-------------------------------------------G-----a-------  2998
CONSENSUS-D   ---------T---G--t-----------------------------------G-----A-------  2995
CONSENSUS-O   ----?--G-----A--A--C--A--?--------------------AC----A--A-----A?-?----  2700
CONSENSUS-U   ----------A-----------------------------------G--------G----  2996
CONSENSUS-CPZ ----G--------G--------A---------------------------G-----A-------  2209
```

HIV1 POL CONSENSUS

```
CONSENSUS-A   GGATGA?GAT                                          2764
CONSENSUS-B   ------G---TAG                                       3011
CONSENSUS-D   ------G---                                          3005
CONSENSUS-O   -ACA--AAG-GAAAGC?TGGAACAGCCT?GTGAAATACCA           2738
CONSENSUS-U   ------G---                                          3006
CONSENSUS-CPZ -A----G---                                          2219
```

HMMER Sequences in the Vif Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_HXB2R	HIVHXB2R	K03455	Starcich,B.	Science 227, 538 (1985)
D_ELI	HIVELI	K03454	Alizon,M.	Cell 46, 63 (1986)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68,1586 (1994)
O_MVP5180	HIVMVP5180	L20571	Gurtler,L.G.	JVI 68, 1581 (1994)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
CPZANT	SIVCPZANT	U42720	Vanden Haesevelde,M.	Virology 221, 346 (1996)
A_ROD	HIV2ROD	M15390	Clavel,F.	Nature 324, 691 (1986)
B_EHOA	HIV2EHOA	U27200	Rey-Cuille,M.A.	Virology 202, 471 (1994)
SD_MM251	SIVMM251	M19499	Franchini,G.	Nature 328, 539 (1994)
STM_STM	SIVSTM	M83293	Novembre,F.J.	Virology 186, 783 (1992)
VER_AGM3	SIVAGM3	M30931	Baier,M.	Virology 176, 216 (1990)
GRI_AGM677	SIVAGM677	M66437	Fomsgaard,A.	Virology 182, 397 (1991)
SAB_SAB1C	SIVSAB1C	U04005	Jin,M.J.	EMBO J. 13, 2935 (1994)
SYK_SYK	SIVSYK	L06042	Hirsch,V.M.	JVI 67, 1517 (1993)

The following alignment was generated using the HMMER program as described in the introduction to this Part and in Part III. For simplicity, only representative types and subtypes are shown. An ordinary consensus sequence (lowercase signifies majority, uppercase signifies 50% or greater) was created from these using MASE; this is not a "most likely sequence" based on an HMM model (Part II). Annotation is based on HIV-1s, therefore the useer should be cautious about its applicability to other PIV sequences.

```
                                                      <- pol end
VIF_CONSENSUS ATGga?aa??????agatgg?t?gtgatg?t?gt?TGGaaggTa?cc?gg??gaggat?aa?a?aTgGa   50
A_U455        -----A--C......------CAG------A-T--A---C-----GA-A--AT--A---T-GA-C-----   64
B_HXB2R       -----A--C......------CAG------A-T--G---C-A---GA-A--AT------T-GA-C-----   64
D_ELI         -----A--C......------CAG------A-T--G---C-A---GA-A--AT------T--A-C-----   64
O_MVP5180     -----A--C......------CAG--AC--A-A--A---C-AA--GA-A-ACA--AAG-G--AGC-----   64
O_ANT70C      -----A--C......------CAG--AC--A-A--A---C-A--GGA-A-ACA--AAG-G--AGCG----   64
CPZGAB        -----A--C......------CAG------A-T--G---C-A---GA-A-AAT------T--G-CT----   64
CPZANT        ---ACTGCA......TCGGTCGGC--C--CGCT--G---C-A---GA-A--AT------T--T-TC----   64
A_ROD         -----GG-AGACAAG------A-A---G-TCCCACC----G---GC-AG--...-----GG-G-A----C   67
B_EHOA        -----GG-GGAAAAG-AT---A-T-CAG--CCGACG----G-A--C-AT-C...---C-AG-G-GG---C   67
SD_MM251      -----GG-GGAAAAG--G---A-A-CAG-TCCCACA----G-A--C-GGA-...---C-AG-G-GG---C   67
STM_STM       -----GG-GGAAAAG--G---A-A---G-TCCCACC----G-A--C-AG--...---C-GG-G-GG---C   67
VER_AGM3      ---A-CC-AGAAAAAGA----G-GA---GAG-AACT--------GC-GGA-GA-CT---A-CT-A----C   70
GRI_AGM677    -----G-GAGAAAAACA----A-A----G-G-A--A----G---GT-AGA-AG-CAA--A-GC-G-----   70
SAB_SAB1C     -----A--G......CAT---A-A--C-G-CCCC-C---------A-GG-CGGACA-CAGG-G-GG----   64
SYK_SYK       -----G--A......GA----A-A---G-TCCCACG----GAA-GA--CCAAGACA---TG-T-G--T-C   64

VIF_CONSENSUS a?ag??TagT?Aaatacca?Atgtata?gac?aaa?a??t????aa????tggtattatagacatcatta  102
A_U455        -C--TT----A---C-T--C------GTCT-C--GA-AGCT...C-AGGT----T------------C--  131
B_HXB2R       -A--TT----A---C----T------GTTT-AGGGA-AGCT...-GGGGA----T------------C--  131
D_ELI         -A--TT----A---C----T------GTTT-A--GA-AGCT...--CAGA----T------------C--  131
O_MVP5180     -C--CC-G--G--------T-AA--C-T-T-T--GA-GGCC...GCGAAC---CG------G--------  131
O_ANT70C      -C--CC----G--------T-A---C-G-T-T-G-A-GACC...G-GAAC----GG--C-----------  131
CPZGAB        -T--TT----T-----T--T--T----G-T-A---A-GGCA...-GAGGT----T---C-----------  131
CPZANT        -ATCCT----G---C----C--A-GGGA---T---GTAC-T...--ACCA---A-G--------------  131
A_ROD         -T--CC-T--C--G--T-TA-AA--C-AA--A--GG-TC-AGAA--GGTG--C---GT-CCC--C---A-  137
B_EHOA        -C--TC--A-T-------TG-AA----G---A---G-CT-GCAAC-GGTC-CT---GT-CCT--C---A-  137
SD_MM251      -T--CC-CA-A-----T-TG-AA----AA--T---G-TC-ACAA--GGTT--C---GTGCCC---TT-A-  137
STM_STM       -T--CC-CA-A---C-TTTA-A-----AC--T---G-GC-AAGC--GGCA--C---GTGCCC--C---A-  137
VER_AGM3      -AG-GA----G-GG---TGG---...-G---T-G-A-AT-A...G-C...---A-A---C--ATG--C--  131
GRI_AGM677    GGG-GA----G-CT---A-G--C...-G--AT---C-AT-G...CCT...---G-A--C----------G  131
SAB_SAB1C     CT--CC----G--------C---C--GT-T-C---C-GTGTGTTC-TTGGC-----ACCCCT---AC-A-  134
SYK_SYK       -GCATA--A-A---ACA--T-A-----AA-GC---G-AT-GGAG--A...GCCACA----A------C--  131

VIF_CONSENSUS ?gaa?g????tgg?a?t??tg?ac??gt????aagta?a?aT?CCacTa??????g?aga?gct?a??Ta  141
A_U455        T---A-T...A-AC-T-CAA-AGTAA--TCAG-----C-C--C------......-GG--A---AGAT--  192
B_HXB2R       T---A-C...CCTC-TCCAA-A-TAA--TCAG-----C-C--C------......-GG--T---AGAT-G  192
D_ELI         T---A-C...CCCC-CCCAAAA-TAA--TCAG-----C-C--C------......-G---A---AGAC-G  192
O_MVP5180     T---TCC...A--A-TCCAAAAGTCA--TCGGCG--GT-T--T---G--......-C---A---G-TA--  192
O_ANT70C      T---TCC...A-AA-TCCAA-AGTTA--TCAAGT---T-T--T---G--......-GG-TG--CC-TG--  192
CPZGAB        T--TCAC...CCTA-TCCAAAAGTTGC-TCTG--A--C-C--A--TT-CAGA...-ATT-TT--A-AT--  195
CPZANT        T---AAT...GATC-CCCAAAG-AAG-AGAAG-----C-T--A------CCAACATTG--TAAAA-AT-G  198
A_ROD         G-TGG-A...---GCA-GG--G--TT-CAGCAGG---ATAT-C---T--AAA...-G-A-CAG-C-TC--  201
B_EHOA        G-T-G-A...---GCT-GG--G--TT-CAGTAG----ATAT-T--C--GAAA...-A--GA--AC-TC--  201
SD_MM251      G-TCG-A...---GCA-GG--G--CT-CAGCAG----ATCT-C--C---CAG...-A--GAAGCC-TT--  201
STM_STM       G-TTG-G...---GCA-GG--G--TT-CAGTAG---GATTT-T--CT-GCAA...-G---A--AC-CC--  201
VER_AGM3      CC--ATTACA---GCA-GG-AC--AATGAGTAG-TATG-G--A--C---GGG...CA-C-T-GAAGTA--  198
GRI_AGM677    GC-GGTGCAA---C-G-TT--G--CTACAGCC-GT-CATT--C--CT--TCA...AA---T-A-T-CA--  198
SAB_SAB1C     AAT-A-A...---A-T-GG-ACT-CTA-CAAG--TGGGTC--A--C--GAAG...-AT-GA---CTAA--  198
SYK_SYK       CC--ATAGAA---C-A-GG-AC--TTA-TGCC--TGGACA--A---G--......-G---T-GAACTA--  195

VIF_CONSENSUS g?agT?a?aa?aTattgg?attTg???acaccaGAaagaggatggct?t??gg?cAtggagt?ag?aTa?  196
A_U455        -T---A-G--C-------GG-C--CAT---GG-----A--AC----AC-TG--T-----G--CTCC---G  262
B_HXB2R       -T-A-A-C--C-------GG-C--CAT---GG--------AC----AT-TG--T--G-----CTCC---G  262
D_ELI         -T-A-A-A--C-------GG-C--CAT---GG--------A-----ATCTG--T--G-----CTCC---G  262
O_MVP5180     -TG--C-CC-C-------GGA--AATGC--GGG-------AGGAA-AC-TG--A-----G--T--T---G  262
O_ANT70C      -T---A-CC-C-------GGA---ATGC--GGG-------ATGAA-AT-TG--A-----G--T--T---G  262
CPZGAB        AT---G-C--CC------GCC--AAGCC--GG------G-CC----AT-TA--C--------CTCC--TC  265
CPZANT        -T---G-C-GTG-TC---GGA--ACAATGTGG----C-CCC-----AC-TA--A-----T--G--C---G  268
A_ROD         -AGA-ACAGGC-------A-C--A...-----------A---------C-CCTCTT--TC---A--A---A  268
B_EHOA        -A---CCA-GG---C---A-CC--...--C--------G----TCT-GAGTTCCT---CT--A--AC--A  268
SD_MM251      -A---ACA-GGG------C-----...---------------G-----CAGTACTT---C---G--G---A  268
STM_STM       -A---TCA-GGG------A----A...-----------A---------CAGT-AAT---CT--G--G---A  268
VER_AGM3      CAT--AGATCT-------C--C--...-----------AG--------A-CAACAT---CT-AGG-G---C  265
GRI_AGM677    -A---G-AT-TT---CACA-CC-C...--C-----------------C-CAA-T--------AG-GT--T  265
SAB_SAB1C     AA---T-CT-AT------C----A...--C-----G-A-------T-AGAAACCT---C-ACAG-A---G  265
SYK_SYK       TGGA-C-CGTTC---CACA-----...G-C---------------T-ACATATG--G---A-T--G--TC  262
```

```
VIF_CONSENSUS aaTgg?????????aagaa??????????Tata??aCagA?gTaga?CCtga???aGCaGAcc???T?aT  236
A_U455        -----AGG...CTG--A-GA..........----GC---C-A-----T-----CCT--------ATC-A--  320
B_HXB2R       -----AGG...AAA----GA..........----GC---C-A-----C-----ACT--------AAC-A--  320
D_ELI         -----AGG...AAA-G--GA..........----GC---C-A-----C----GCCTG-------AAC-A--  320
O_MVP5180     -----CAA...TAC---G-G..............-----AA---C-GA-T--T-----AAC-------AGGA-G--  320
O_ANT70C      -----CGT...TAT-----G..............-----AA---C-GA-T--C-----AAC-------AGGA-G--  320
CPZGAB        -----AGG...CTAGGATCT.............---GTC---C-G--G--T--CTTCAC------TAGAT-G--  323
CPZANT        -----AGA...TGTGGA--G.............----TC---C-A-----C--A--AAC---T--T-AAA-G--  326
A_ROD         CT---TAC...ACAG-A--G.........-TCTGG-----T--TACC--A--CTGT-----TGTCC-A--  326
B_EHOA        C----TAT...GAG-G--GC.........-T-TAT-----T---ACT-----TGT--------GAT-AC-  326
SD_MM251      CC---TAC...TCA-G---C.........-T-TGG-----T---ACA--A--CTAT------ATTT-AC-  326
STM_STM       CT---TAT...ACC-GA--T.........-TCTGGT----T---ACA--A--CTGT-----T-AGC-AC-  326
VER_AGM3      -G-ATCTA...AGC--T-GGGATCCTTGG----GG-----AT-G--T----CAAC------TAGCC-G--  332
GRI_AGM677    CC-ATTACCATCAA---GGA..........----AG-----A-----T--A-GAAC-------AGAA-G--  326
SAB_SAB1C     G--ACAGT...AAAGG-G-A.........-GGTTT-----GC-T--T---TGGAC---G----ATA-T--  323
SYK_SYK       ---ACCAA...TGG--CC-G.........-GG-AC--T--CT--ACA--A-CAGT--------GCC-T--  320

                                          \/3'sj
VIF_CONSENSUS ?CAt????atTAtTTc?ctTgtTTTaCAGaa??aGc?gTcagaaaaGCcaTc?taGGagAaa?a?t??tg  292
A_U455        T--CCTGC-------TGAC------T-----TCT--CA-A--G-G------AT-----C---T-G-TAGA  390
B_HXB2R       T---CTGT----C--TGAC------T----CTCT--TA-A-----G---T-AT-----C-C-T-G-TAGC  390
D_ELI         T---ATGT-------TGA-------T-----TCT--TA-A-----------AT-------T-T-G-TAGT  390
O_MVP5180     A---CTGC--------A-A-----------TC---AA----G--G-----TC----GC-G-G-G-GC--  390
O_ANT70C      A---CTGC-------TA-C----------C-TC---A-----G---------C-----C-G-G-G-AC--  390
CPZGAB        C--CTCTC-A-----TGA---C---G-----ACT--CA------G---T--AT----GC-GCTTG-AGC-  393
CPZANT        A--CCAAT--------C-C------T----TCA---A------C----A---C---------G-A-AT--  396
A_ROD         A---AGCAC-------C----C-------C-GGT-AA--A----G-------AG---G-----AGT-AT--  396
B_EHOA        G---GGGTC-------T---CG-------CTAAT-AA--A--G-G-------AGG-------AGA-AT--  396
SD_MM251      G---AGCAC-------C----C-------CGGG--AA--G----GG------AGG------CA-C-GC--  396
STM_STM       G---GGCAC-------C----C-------CGGG--AA--G----G-------AGG-------AGT-AC--  396
VER_AGM3      A---ACCC--------TA-----------AGG--CA----G------C-AT-G---C-G-GGT-CACC  402
GRI_AGM677    A--CCTAT-------TAAC----------TAG---CA--CA-C-G--T---AG---G--G-AGTATAC-  396
SAB_SAB1C     T---TGGTCC------C-C----------TAG---T---CA-C--------AGG-------AGTATC-T  393
SYK_SYK       A---AATTT-------C----C-------C-AG---G----ACC-G--AG--AG---G---CT-T-GACA  390

                                                    5'sj\/
VIF_CONSENSUS tc????TGt?a?tac????aagcaggacat?????????agccAgGTAc?atc?cTaCAgtattTaGC?c  340
A_U455        C-TAGG---G-A--T...C----------.........-A-A-----GG---T-----A-----G--T-  448
B_HXB2R       C-TAGG---G-A--T...C----------.........-A-A-----GG---T-----A--C--G--A-  448
D_ELI         C-TAGG---G-G--T...C----------.........-A-A-----GG---C----------G--A-  448
O_MVP5180     A-CAAG---G-A---...CTG--------.........--T------GGGA-A-----A-TC-----CT  448
O_ANT70C      A-CAAG---G-A---...CCTA-------.........--T------GGGA-A-----ACTGC----T-  448
CPZGAB        C-TCGT---G-A---...A---A------C.........--G------GG---CT------TC--G--C-  451
CPZANT        A-ATAC--CC-C---...A-GAAG-----.........--T------GGGA-A--G----------TT  454
A_ROD         --CTGC--CA-T--TCCCCG---TCATAGA.........GC--------CG--A--T--A-T-C-G--CT  457
B_EHOA        --CCAC--CA-C---CCATC---TCATACA.........G---------C-AGTT------T-C----C-  457
SD_MM251      --TTGC--CA-G-T-CCGAG---TCATAGG.........TA--------C-AGC--------C-----A-  457
STM_STM       --TTGC--CAGG-T-ACGA----TCATA-G.........-A--------C-AGCT-------C-----TT  457
VER_AGM3      -TC...---C-G-T-CCCG-G-G-CACA-GAAA...ACAG-A-------CC--TT-G--A--C-----T-  466
GRI_AGM677    -GG...--CACA-T-AAGG---G-CATA-A.........G-T-------A---A--G--ACT---G--A-  454
SAB_SAB1C     -GG...---A-AC--...C-G-TG------CAGCCAACAG-G--A----C---AT----------A-  457
SYK_SYK       --ACAC--CTGGAC-CCA......CATAC-.........GA--------CC--A--T----------AT  445

VIF_CONSENSUS Taa?agca?????a???????ta?????????a?a?ccaag????????????ag????aa?c?ac???t  365
A_U455        -G-A----...TT-......G--.........-CCC-A-CA..........--GGCA--G-C--CTT-  488
B_HXB2R       --GC----...TT-......A--.........-C-C-A--A..........-AGATA--G-C--CTT-  488
D_ELI         ---C----...TT-......A--.........GC-C-A--A..........CAGATA--G-C--CTT-  488
O_MVP5180     -G-A----GTA.........G-G.........-A-GTA--A..........--AAAT--G-CT-CCC-  488
O_ANT70C      ---G----GTA.........G--.........-A-G-A-GA..........--CAGG--G-CT-CCC-  488
CPZGAB        ---A---T...TT-......A--.........TC-GAG-GA..........--GCAT-GA-C--CTT-  491
CPZANT        -TTGTAAGATCCT-GAA...T-CAGA...GGATACC-----..........G-TCCA-GAAG--AGT-  506
A_ROD         --GTG-T-GTGCA-......CA-AAT...GAC-G-C--C--AGAGACAGTACC-CCAGG--A-AG-GGCG  518
B_EHOA        ---G--TTGTACA-GAA...GG-AAA...GATGG-T--C--GGAGAGAGTACC-CCAGG--A-AG-GGCG  521
SD_MM251      ---A--T-GTAAGCGAT...G-C.........-G-T--C--GGAGAGAATCCC-CCTGG--A-AGTGGAG  515
STM_STM       ---A--T-GTGGA-CAT...G-C.........-G-T--C--AGAGAGAATACCGCCCGG--A-AGTGGAG  515
VER_AGM3      -CCTT---CACCA-AAT...GGCCTCAGGCAG-G-T--C--AGAAGCAAGACCG-GGGA-CTAG-AATA-  533
GRI_AGM677    --GTT---TATAC-AATGGCA-CAGG...AAG-G-T-----AGAACCTTTACC--GATGGCTGG-AATC-  521
SAB_SAB1C     -T-GG-T-TATAC-AATGGCC-C.........-GGAGGGT-..........GCTCCC-CCAGTAGGAG  506
SYK_SYK       --CA--TGTACCT-......AA-.........GATGG-GGA..........G-CTTTCTT-A-TCCC-  488
```

```
VIF_CONSENSUS ?ccaagt?????????????????????????????????????????????????????????????????????  371
A_U455        G--T---.....................................................................  495
B_HXB2R       G--T---.....................................................................  495
D_ELI         G--T---.....................................................................  495
O_MVP5180     A--C---.....................................................................  495
O_ANT70C      A--C---.....................................................................  495
CPZGAB        A--T---.....................................................................  498
CPZANT        C--T---.....................................................................  513
A_ROD         AAG-GAC.....................................................................  525
B_EHOA        AAG--AC.....................................................................  528
SD_MM251      AAG-GAC.....................................................................  522
STM_STM       AAG-G-C.....................................................................  522
VER_AGM3      GGG-TT-.....................................................................  540
GRI_AGM677    GGG-TC-.....................................................................  528
SAB_SAB1C     GGG-TCCTCCCAAGGATCCCCCCAAGAATCCCAGAGAAGAGATACCAGGATGGCTAGAAACATGGGATTT  576
SYK_SYK       C---GCA.....................................................................  495

                                                                          vpR ->
VIF_CONSENSUS ???????????gtt??g???????????aga?tg???acaga?gat??tagatg?aa?aAa????g??  400
A_U455        ...........---AA-...........-A-T-A..-----G---..-----G--C--G.......  528
B_HXB2R       ...........---AC-...........-A-C--..-----G---..-----G--C--G.......  528
D_ELI         ...........---AG-...........-AGC-A..-----A---..-----G--C--G.......  528
O_MVP5180     ...........--CCA-...........---T-A..-----A---..-----G--C--G.......  528
O_ANT70C      ...........--CCA-...........-A-T-A..-----A---..-----G--C--G.......  528
CPZGAB        ...........--CGCC...........-A-T-A..-----G---..-----G--C--G.......  531
CPZANT        ...........C-GTCT...........-TCT-A..-----A---..-----G--C--G.......  546
A_ROD         ...........TA-CG-AGAGGC...CTTC-CC--GCT-A-C-G--CAG----A-CC-T---CAGA-AA  580
B_EHOA        ...........AG-AG-AGAGGC...ATTC--A--GCT-G---CA-CAT----ACATCTC--CAGA-TA  583
SD_MM251      ...........AA-AG-AGAGGC...CTTC--A--GCT-A-C-GA-CAG----G-AG-T---CAGA-AG  577
STM_STM       ...........AA-AGAGGAAGC...ATTC--G--GCT---C-GA-CGG----G-AC-T---CCGA-AG  577
VER_AGM3      ...........-AGCA-GGAGCCGTGGGAC--A--GCT-AGAGAC--GC----A-AT-TC--TCAG-AA  598
GRI_AGM677    ...........AGACA-GGAGCCATGGGAC--A--GCT---AGAC--GC-CAGG-ATCT---CGAA-AA  586
SAB_SAB1C     GCCCAGAGAGCC---CGA...........C--A--GCTC-GAGAC--GT--C-G-ACCTC--TTCA-AG  634
SYK_SYK       ...........TG-GCA...........--GAAC...--CATG-T-CT-CAT-CC--G---TGTA-AG  538

VIF_CONSENSUS ???????????????????????????????????????????????????????????????????????  400
A_U455        .....................................................................  528
B_HXB2R       .....................................................................  528
D_ELI         .....................................................................  528
O_MVP5180     .....................................................................  528
O_ANT70C      .....................................................................  528
CPZGAB        .....................................................................  531
CPZANT        .....................................................................  546
A_ROD         GC...................................................................  582
B_EHOA        GC...................................................................  585
SD_MM251      GC...................................................................  579
STM_STM       GC...................................................................  579
VER_AGM3      GCCAAGATGCATTTTGGGCG.................................................  618
GRI_AGM677    GCCAGAAGGCACTT.......................................................  600
SAB_SAB1C     GCCCAGTGCCACTTCCC....................................................  651
SYK_SYK       TTGACCCCAAACGTGATCAATGTCACTGTAAAGGCAGAACTGGTAGTGACAGAAGCATCCAAGCATTTTA  608

VIF_CONSENSUS ?????????????????????????????cc?c??a?ga??ag?g???a?aga???ggg??????agccat  422
A_U455        ............................--C-AG-A--CC--G-GCC-C---...---.....----G-  561
B_HXB2R       ............................--C-AG-A--CC-AG-GCC-C---...---.....-----C  561
D_ELI         ............................--C-AGCA--CC--G-GCC-C---...---.....------  561
O_MVP5180     ............................--CTGG-AA-TC--G-ACC-GCT-...---.....------  561
O_ANT70C      ...........................-AC-TG-GA-TC--G-ACC-GCT-...AA-.....----C-  561
CPZGAB        ...........................-AC-AG-G--CC-AG-TCC-CCA-...-A-.....-A--T-  564
CPZANT        ............................--C-GC-G--TG--G-GCC-C---...-A-.....-A---A  579
A_ROD         .....................AGTGAAT-A-CT-CCCCG--AACTT-TTTTCCA--TGTG...GCAG-G  627
B_EHOA        .....................AGTCAAT-A-TTGCCCAGG-AACTT-TTTTCCA--TCTG...GCAG-A  630
SD_MM251      .....................AGTAAA--A-CT-CC-AGG-A-CTG-TTTTCCA--TTTG...GCAA-G  624
STM_STM       .....................AGTAAA--ATCT-CCGAGG-AACTG-TTTTCCA--TCTG...GCGA-G  624
VER_AGM3      .......CGAGCTCCTGTTCCAAGTATGGAA-TATT-TCAG-A-GAGGG---AAG-AATCGCACTC----  681
GRI_AGM677    ..........TGGAATGAACATGCTAAT--GAGT-T-GAATTACTGTGT---...--A.....G-GA-G  651
SAB_SAB1C     ..........AAGGAATCTCCTTTTCCGT-TTTGGT-GAATATTGTGGA---...ACC.....------  702
SYK_SYK       CTCCTCAAGAAATATATGGAGTTTGGAAT-AATCCTT-AAC-AAGAGGC--GGAC              663
```

```
VIF_CONSENSUS ????c?a?gaat??????????????gggca?  433
A_U455        ...A-GCT----.............A----T  576
B_HXB2R       ...A-A-T----.............--A--C  576
D_ELI         ...A-A-T----.............-----T  576
O_MVP5180     ...T-A-T----.............--A--C  576
O_ANT70C      ...T-A-T----.............-----C  576
CPZGAB        ...A-C-G----.............-----T  579
A_ROD         GTCCTGGA--TA.............CT-GCA  645
B_EHOA        GTCTTGGGC-TA.............TT-GCG  648
SD_MM251      GTCTTGGGA-TA.............CT-GCA  642
STM_STM       GTCTTGGGA-TA.............CT-GCA  642
VER_AGM3      ...G-T-GA--G.............--CTTA  696
GRI_AGM677    GAGA-A                          657
SAB_SAB1C     TGAT-A......TGGGCAAACCAGGCT--AAGG  729
```

HIV1 VIF CONSENSUS

```
                                                                                                  <- pol cds
CONSENSUS-A   ATGGAAAACAGATGGCAGGTGATGATTGTgTGGCAGGTAGACAGGATGAgGATTAGAACATGGAAcAGTT      70
CONSENSUS-B   ------------------------------------a-----------------------------a----      70
CONSENSUS-D   ------------------------------------A-----------------?---------A----      69
CONSENSUS-O   --------------------AC----A--A-----A?-?-----ACA--AAG-G-A-G-?--------CC      67
CONSENSUS-CPZ ---?????????????????--?--???---------A--------?----------A?-??-----????-      43

CONSENSUS-A   TAGTAAAACATCAtATGTATGTCTC?AAGAAAGCTaAgGGTTGGTTTTATAGACATCACTtTGAAaGTAG     139
CONSENSUS-B   ----------C---------a-t--Agg-----------a------------------A------c-c     140
CONSENSUS-D   ----------C-----------T--A----------?CA-A------------------A---c--CCc     138
CONSENSUS-O   -?--G---T-C----A?--CA?G--T-??--G?-CG?-AAC---?G?--?--?-----T-A----TCC--     126
CONSENSUS-CPZ ----?---?-?--?--?-????G?----A??????-?A???---???--?--------T-A---??A???      83

CONSENSUS-A   gCATcCAAaAGtAaGTTCAGAAGTACACATCCCACTAGGGGAt......GCTAGaTTAGTAGTAAGAACA     203
CONSENSUS-B   T-------g-a-----------------------------.....--------g---a---c----     204
CONSENSUS-D   C--C------A------------------?--------A--A.....------C-G-------a----     201
CONSENSUS-O   ?A------?---?-----????--?T-T--T---G---??-??.....--??AT?----?--?-CC---     173
CONSENSUS-CPZ T?-?-----????G?????---?----?--A--??-???A???T????????-A---??----G-C-???     119

CONSENSUS-A   TATTGGGGTCTGCATACAGGAGAAagAGACTGGCACTTGGGTCAtGGGGTCTCCATAGAATGGAgGCaGA     273
CONSENSUS-B   ----------------------------------t--------g--A------------------A-a-     274
CONSENSUS-D   ----------------------?-----TC-------G--A------------------A-A-     270
CONSENSUS-O   --------AT-?ATGC----G---------?GAA--?-----A--------TAGT---------C??T-?-     237
CONSENSUS-CPZ -??----??T-A??????------?-??C?-----?--A--?-----?--???---??-------????G     163

CONSENSUS-A   AAAGATATAGCACACAAgTAGAtCCTGACcTAGCAGACCAaCTgATTCACCTGcATTATTTTgAcTGTTT     343
CONSENSUS-B   -G--------------------c--------------------A-----t---t----c-----t-----     344
CONSENSUS-D   GG--------------------C----G---G-----------A-----TA--T----------T-----     340
CONSENSUS-O   -G?AG----AA-----GA-T--?-----AAC-------AGGA----A--T----------?AC?-----     303
CONSENSUS-CPZ G-???---?T------?--?--?--????AC---?--T??-?----?---????-?-----????--?--     207

                \/ 3'sj
CONSENSUS-A   TTCAGAcTCTGCCATAAGgAaAGCCATATTAGGAgAAaTAGTTAGACCTAGGTGTGAATATCAAGCAGGA     413
CONSENSUS-B   ------a-----t-----a--t----------c-t--------t----------------------     414
CONSENSUS-D   ------A?----?-----A---------------c-T-----?--T-?---------G----------     406
CONSENSUS-O   -A---?A--A--A?-C-----?-----?C----?C-G-G---?CTGA-C-A---------C-???-----     364
CONSENSUS-CPZ -?----????--??-C--A??---?--??----??-?????-A??G?-????--??-?--CA-??A?---     245

              5'sj \/
CONSENSUS-A   CATAAcAAGGTAGGATC?CTaCAATATTTGGCaCTG...AAaGCaTTA.........GTAgCaCCAAcAA     470
CONSENSUS-B   ----------------t-----g--C--------A...gC-------.........a--a------a--     472
CONSENSUS-D   ----?------------Ct----G-----------A...-C-------.........a---------A--     463
CONSENSUS-O   ----GTC-------GA-A------?T??-A--??-?...-?----G--.........--?AA-G?--?--     412
CONSENSUS-CPZ --?-G?C-------??--?-?--G-??--?--??-????--?????--???????????-T?????-???     275

                                                                    vpR cds ->
CONSENSUS-A   aGgCAAAGCCACCTTTGCCTAGTgTTAaGAAgTTAaCAGAaGAtAGATGGAAcgAGCCCCAGAAGACCAG     540
CONSENSUS-B   --At----------------------?---AC-g-----g------------A---------------A     541
CONSENSUS-D   --AT------?----------------G----?--------------------A---------------a     531
CONSENSUS-O   G?A??-----T--CC-A--C----CC---?A---------------------A---?-??--?A-T---     474
CONSENSUS-CPZ G?C??-GA??--??--?------?-??C?-??--------?------------A---?--??-G--??-?     324

CONSENSUS-A   GGGCCACAGAGGGAgCCGT?CaaTGAATgGaCActAG                                     576
CONSENSUS-B   -----------------a-a--------------?--                                     577
CONSENSUS-D   --------?--------A-A-------------T                                       564
CONSENSUS-O   --A---GCT-??-----?-T----------?---?--                                     506
CONSENSUS-CPZ --?----??--A--A--??--C-G------G--T                                        352
```

HMMER Sequences in the Vpr Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_HXB2R	HIVHXB2R	K03455	Starcich,B.	Science 227, 538 (1985)
D_ELI	HIVELI	K03454	Alizon,M.	Cell 46, 63 (1986)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68, 1586 (1994)
O_MVP5180	HIVMVP5180	L20571	Gurtler,L.G.	JVI 68, 1581 (1994)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
CPZANT	SIVCPZANT	U42720	Vanden Haesevelde,M.	Virology 221, 346 (1996)
VER_AGM3	SIVAGM3	M30931	Baier,M.	Virology 176, 216 (1990)
GRI_AGM677	SIVAGM677	M66437	Fomsgaard,A.	Virology 182, 397 (1991)
SAB_SAB1C	SIVSAB1C	U04005	Jin,M.J.	EMBO J. 13, 2935 (1994)

The following alignment was generated using the HMMER program as described in the introduction to this Part and in Part III. For simplicity, only representative types and subtypes are shown. An ordinary consensus sequence (lowercase signifies majority, uppercase signifies 50% or greater) was created from these sequences using MASE; this is not a "most likely sequence" based on an HMM model (Part II). Annotation is not attempted for this alignment of paralogous sequences.

```
VPR_VPX_CONSENSUS ATGGaacaAGcc??????????????????????????????????????????????????Cca???g     16
A_U455            ------------.....................................................---...-     16
B_HXB2R           ------------.....................................................---...-     16
D_ELI             ------------.....................................................---...-     16
O_ANT70C          ----------A.....................................................--T...-     16
O_MVP5180         ------------.....................................................-TG...-     16
CPZGAB            ----------A.....................................................---...-     16
CPZANT            ------------.....................................................--G...C     16
VER_AGM3          ----CCTC--G-AGA.....................................................GAT--CAGA-  25
GRI_AGM677        ----C-TC--GAAGA.....................................................GAT---AGA-  25
SAB_SAB1C         ----CCTC--GAGGGTGGCTCCCACCAGTAGGAGGGGATCCTCCCAAGGATCCCCCCAAGAAT--CAGA-     70

VPR_VPX_CONSENSUS aa???...ga?cagGGg???????????????????ccaca?ÀggGAgCCat?c...aatGAaTgG?ca..     53
A_U455            --.......--C------.....................-----G--------G-A-...GC-----A-G--..     57
B_HXB2R           --.......--C--A---.....................-----G---------CA-...---------A--..     57
D_ELI             C-.......--C------.....................-----G---------A-...---------G--..     57
O_ANT70C          -G......A-T-----A.....................---GCT-AA-----C-T-...---------G--..     57
O_MVP5180         --.......A-T-----A.....................---GCT----------T-...---------A--..     57
CPZGAB            -G......--C--A--T.....................----CA--A--A--T-A-...C-G------G--..     57
CPZANT            -G......--TG-----.....................-----G--A--A---ATG...-----G---TTG..     57
VER_AGM3          --GCA...AGA-C----GAACTAGAAATATGGGATTTGAGC--------G-GG...G-C------CT-..     87
GRI_AGM677        --CCT...TTA-CA--ATGGCTGGAAATCTGGGAT-T-G-C----------GG...G-C------CT-..     87
SAB_SAB1C         --GAG...ATA-CA--ATGGCTAGAAACATGGGATTTG-CC--A-----G-T-...G-C------CTC..    132

VPR_VPX_CONSENSUS ....ctA...GAgat?tTagaaGAgcT?AA??atGAaGc?gtaagaCAtTT?...cctaggc?ttggcT?    105
A_U455            ....T--...---T-G-----------T--GC------AT--T--------T...-------AG-----C    117
B_HXB2R           ....---...---C-T-----G-----T--GA-------T--T--------T...------AT------C    117
D_ELI             ....T--...---C-T-----G-----T--GAG------T--T--------T...------ATA-----C    117
O_ANT70C          ....---...---C-C-----------A--AGCA-----A----------C...-------C------A    117
O_MVP5180         ....---...---C-CC-G--------G--AG-A-----A----------C...-------C----T-A    117
CPZGAB            ....T--...----CC--G-----AT-A--AA-------A----------C...-----A-CA---T-G    117
CPZANT            ....T--...--A-CT--------AA-A--AA----G--A----------C...--ACA--C-ACAT-A    117
VER_AGM3          ....AG-...--C--GC-------TA-C--TC-G-----CAAG-TG-----T...GGGC-CGAGCTC--G    147
GRI_AGM677        ....-A-...--C--GC-CAGG--T--A--CG-A-----CAG---G--C--T...GGA-T-AACAT---A    147
SAB_SAB1C         ....-G-...--C--G---C-G--C--C--TTCA--G--CCAGT-C--C--C...--A---AA-CTC--T    192

VPR_VPX_CONSENSUS ca?cgcttagg??????????caatatatttatgAg????????????actttatggggatac???????t    145
A_U455            --TG-A-----A.........---C----C---A-C.............--C----------T......-    160
B_HXB2R           --TG-------G.........---C----C-----A.............-------------T......-    160
D_ELI             --TA-------A.........---C----------A.............-------------C......-    160
O_ANT70C          --CGC---G--A.........-----C---------.............----------C--T......-    160
O_MVP5180         --AGC--GT--G.........--G--C---------.............--------A--C--T......-    160
CPZGAB            --T-AA-----A.........----TC--------C.............--------A-----A......-    160
CPZANT            --A-A---G-AAATTGGGTATATGC-A---ACG---.............-T-C----A--GGAG.......-    169
VER_AGM3          TTC-AAG--T-G.........A-C---TG-C-G---GAGGGA...GAA-GGA--C-CAC-C-CATG...C    202
GRI_AGM677        ATC--AG--T-G.........A-T--CTG-GTA---GAGGGA......-GGAGACATA----CCCATGGA    202
SAB_SAB1C         TTC--TC-TT-G.........TGGA-----GTG--AGAACCAGCCATTGA-C-----C-A--CAGG...C    250

VPR_VPX_CONSENSUS gggaaggagtt.............?a?gc?at?ataAgAaT??Tg...CAAcAatt?cTgTTTat?CAT??    190
A_U455            ----G------.............G-A--T--A--------TT--...--------A--------T---..    213
B_HXB2R           ---C------G.............G-A--C--A--------TC--...------C-G--------C---%%    213
D_ELI             ---T-------.............G-A--T--A--------AC--...--------A--------T---..    213
O_ANT70C          ---T-------.............ATG--A--T--------CT-A...--------G-------CC---..    213
O_MVP5180         -----------.............ATG--A--T--------CT-A...------C-A-------CC---..    213
CPZGAB            ---T------A.............G-G--C--C--------CT-A...-----C--G--------C---..    213
CPZANT            -CAGGTTT                                                                  177
VER_AGM3          TA---A-G-C-.............T-TAAATATTAT-A-T-GG--...---A--GCT--C---G-G---..    255
GRI_AGM677        AT--GAT--GC.............T-CAAGTACTAT-----TG-T...---A-G-CTA-----G-A---..    255
SAB_SAB1C         T------CTGG.............T-TAAATATTGT-----TC--...---A-GGCTT-A---G-G---..    303
Z
VPR_VPX_CONSENSUS ?ttcagaatTgG?TGccaacatAg??????????????????????????????????????aGaatagg?    220
A_U455            .----------G----------C.....................................--------C    246
B_HXB2R           T----------G--T-G---                                                       234
D_ELI             .----------G--T--------C.....................................--------C    246
O_ANT70C          .-AT--------A----------T.....................................--------A    246
O_MVP5180         .-AT--------A----------T.....................................--------A    246
CPZGAB            .--T---C----G-----G-----C.....................................--------C    246
VER_AGM3          .---C-GTG---A----GCAGA--ACAACCCTTTGAACCA.........TACGAAGAAAGG--GGAT--A    315
GRI_AGM677        .------TG---T--TAG-AGG--AGGACCTTTTTCCCCT.........TACGAAGAGAGG----AT--A    315
SAB_SAB1C         .A-G-A-GG-A-A--T...TG--AGCCAAAGACTCACCCTGCGTATGGACCAGGAGCAGGAG-CCCGCCC    369
```

```
VPR_VPX_CONSENSUS att????????????a??????gAgga...aga?gA????????ggaag?aatggatCcagtagatcc  257
A_U455            ---............-TT...CC---G...---A--.........--C--G------G----------  288
D_ELI             ---............-TT...C--CAG...---A--.........-C---A-----------------  288
O_ANT70C          ---............-ACCCAA------...---G--.........A----A-----------------  291
O_MVP5180         ---CTCCCATCTAAC-CA...A------...---G--.........A----A-----------------  300
CPZGAB            --C............CTC...CC-CA-...---A--.........A--TCC----------A-------  288
VER_AGM3          CAA............GGG...G-------...C--GC-.........--GC-CGTACC-C-AG-ACTTGAT 357
GRI_AGM677        CAA............GGA...G------...GCCCC-............CCCCC-CCTC-AG-ACTT    351
SAB_SAB1C         CCA............GGTCTGG------...GCTTC-GGAGGAGCT-C-TCGGCC-CTC-AG-CCTG    420
```

HIV1 VPR CONSENSUS

```
                                                                                                                <- vif cds
CONSENSUS-A   ATGGAA??A......GCCCCA...GAAGACCAGGGGCCACAGAGGGAGCCGT?C??TGAAT?G?CA?TAG                                 53
CONSENSUS-B   ------ca-??????------???--------A-----------------a-A-AA-----G-a--C---                                 61
CONSENSUS-D   ------CA-......------...--------a---------?--------A-A-AA-----G-a--T---                                 60
CONSENSUS-O   ------CA-.......--?-??...--?A-T-----A---GCT-??-----?-T-AA-----G----C---                                 53
CONSENSUS-U   ------CA-......-------...-C------A-----------------ACA-AA-----G-A--T---                                 61
CONSENSUS-CPZ ------CA-.......--?--?...?-G--??-?--?----??--A--A--??-?-A?--?-G--??T---                                 42

CONSENSUS-A   A??TGTTAGAAGAGCTTAA?CATGAAG?TGTTAGACA?TTTCCTAGGC?GTGGCTCCATGGATTAGGACA                                  117
CONSENSUS-B   -GC-T-----g--------GAg-----C---------T---------ata---------a-c--------                                  131
CONSENSUS-D   -GC-T-----G--------GAG-----C---------T---------ATA---------A-C--------                                  130
CONSENSUS-O   -GC-C?-?--------?--AG?A----CA--A-----T--C-------CT---?-A--?-CC-??--?--                                  114
CONSENSUS-U   -AC-T-----G--------G--A----C---C-----C---------ATA---------A-T--------                                  131
CONSENSUS-CPZ --AC?--?-----A?-A--AA----?-CA--A-----T--C--????-C????T-?????????????--                                   86

CONSENSUS-A   ACATATCTAT?A?AC?TATGGGGATACTTGGGA?GG?GTT?AAGCTATAATAAGAATTTTGCAACAA?TA                                  180
CONSENSUS-B   ----------G-a--T----------------cA--A--gG----c----------c--------C-G                                    201
CONSENSUS-D   -?----t---G-A--T----------C----?A--a---G-----?----------?--------t--                                    196
CONSENSUS-O   ?T-C--T---G-G--T-----?--C-------?A--A---ATG--A--T--------C--A--------?                                  179
CONSENSUS-U   ----------G-A--T----------C-----A--A---G---------------G-C-.-------C-G                                  201
CONSENSUS-CPZ -??C?-??-?-?-??----??????????????-??-???-?????????????????-?---???--?                                   98

                              \/ 3' sj
CONSENSUS-A   ?TGTTT?TTCAT...TTCAGAATTGGGTG?CAACATAGCAGAATAGGCATTATT.........C?AGGGA                                  234
CONSENSUS-B   C-----A------..?--------------t-g--------------------c-.........---Ca--                                  258
CONSENSUS-D   C-----A------...--------------T-----------------t----c-.........-G-CA--                                  254
CONSENSUS-O   C-----ACC---...-AT--------A--C--------T--------A---??C??????????C-A-AG                                  234
CONSENSUS-U   C-----A------...--------------T----------------------C-.........-G-CA--                                  259
CONSENSUS-CPZ -?-???A-?---...????--???-?---C??G??---------------CC-C.........-C-CAA-                                  138

                           -> tat cds
CONSENSUS-A   GAAGAGGC...AGGAATGGAGC??GTAGATCCTAG                                                                    264
CONSENSUS-B   -g----CA???--a--------CA-----------                                                                     290
CONSENSUS-D   ------CA...--A------T-CA--------                                                                        283
CONSENSUS-O   -------AAGA--A------T-CA--------                                                                        266
CONSENSUS-U   ------CA...--A------T-CA--------                                                                        288
CONSENSUS-CPZ -----A-A...TCC------T-CAA-------                                                                        167
```

HMMER Sequences in the Tat Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_HXB2R	HIVHXB2R	K03455	Starcich,B.	Science 227, 484 (1985)
C_UG268A2	HIVUG268A2	L22948	Louwagie,J.J.	JVI69, 263 (1995)
D_ELI	HIVELI	K03454	Alizon,M.	Cell 46, 63 (1986)
F_BZ163A	HIV1BZ163A	L22085	Louwagie,J.J.	ARHR 10, 561 (1994)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68, 1586 (1994)
O_MVP5180	HIVMVP5180	L20571	Gurtler,L.G.	JVI 68, 1581 (1994)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
CPZANT	SIVCPZANT	U42720	Vanden Haesevelde,M.	Virology 221, 346 (1996)
A_ROD	HIV2ROD	M15390	Clavel,F.	Nature 324, 691 (1986)
B_EHOA	HIV2EHOA	U27200	Rey-Cuille,M.A.	Virology 202, 471 (1994)
SD_MM251	SIVMM251	M19499	Franchini,G.	Nature 328, 539 (1994)
STM_STM	SIVSTM	M83293	Novembre,F.J.	Virology 186, 783 (1992)
VER_AGM3	SIVAGM3	M30931	Baier,M.	Virology 176, 216 (1990)
GRI_AGM677	SIVAGM677	M66437	Fomsgaard,A.	Virology 182, 397 (1991)
SAB_SAB1C	SIVSAB1C	U04005	Jin,M.J.	EMBO J. 13, 2935 (1994)
SYK_SYK	SIVSYK	L06042	Hirsch,V.M.	JVI 67, 1517 (1993)

The following alignment was generated using the HMMER program as described in the introduction to this Part and in Part III. For simplicity, only representative types and subtypes are shown. An ordinary consensus sequence (lowercase signifies majority, uppercase signifies 50% or better) was created from these representative sequences using MASE; this is not a "most likely sequence" generated from a HMMER model (Part II). Annotation is based on HIV-1s, therefore the user should be cautious about its applicability to other PIV sequences. Putative splice sites are designated by '\/'.

```
                                  <- vpR end
TAT_CONSENSUS ATGga?cca?taga?cc?gag?????????????tagagccctggaacca?cc?????????????????      35
A_U455        -----G---G----T--TA-C............C-------------A--C--GGGAAGT.........      48
B_HXB2R       -----G---G----T--TAGA...........C-------------G--T--AGGAAGT.........      48
C_UG268A2                                                    %-AGGAAGT.........       8
D_ELI         -----T---G----T--TA-C............C---------------T--AGGAAGT.........      48
F_BZ163A      -----G-T-G----T--TA-C............T----T-----------T--AGGAAGC.........      48
O_ANT70C      -----T---G----T--T---............G-GCCC--T---C----C--TGGAAGT.........      48
O_MVP5180     -----T---G----T--T---............A-GCCC--T---C-T--C--TGGGAGC.........      48
CPZGAB        -----T---A----T--T--C............C-G-----------A--C--AGGCAGT.........      48
CPZANT        -----C--CG----CG-T--A............AC-CCT--A---TTA--C--TCCTGCT.........      48
A_ROD         -----GA--CCCTTGAAG-C-CCAGAGAGCTCAT--A--T----C---G-G--CTTTTCACGCACTTCAG      70
B_EHOA        -----AAT-CCCTTGAAG---CAGGAGAGCTCAT--A-CT---CC-G-GGG-ACTCTTCCTCCACTTCAG      70
SD_MM251      -----GA--CCCTTGAGG---CAGGAGAACTCAT----AT---CC---G-G-GCTCTTCATGCATTTTAG      70
STM_STM       -----GA--CCCTTGAAG---CAGGAGAGCTCAT--AGAT---CC-G-G-G--CTCTTCATGCACTTCAG      70
VER_AGM3      ---.........~~CAAG-G-............GAG--CGAGCA-GG-GCGTACCACCAG.........      39
GRI_AGM677    ---.........~~CAAG---............GAG------.....--C--CCTCCTC.........      33
SAB_SAB1C     ---.........~~C-AG---CAGGAGGCCCGCCCCC--GT----G-GG-G-TTCAGGAG.........      51
SYK_SYK       ---TCCT--ACG--C-AGATA............TGCC--A-ACA--GGGTA--CCCATCCTTCTTAGAAG      58

TAT_CONSENSUS ??????????????????????????????????????????????????????????????????CC      37
A_U455        ..........................................................CAG...--      53
B_HXB2R       ..........................................................CAG...--      53
C_UG268A2     ..........................................................CAG...--      13
D_ELI         ..........................................................CAG...--      53
F_BZ163A      ..........................................................CAG...--      53
O_ANT70C      ..........................................................CAG...--      53
O_MVP5180     ..........................................................AAG...--      53
CPZGAB        ..........................................................CAG...--      53
CPZANT        ..........................................................ACA...--      53
A_ROD         AGCAGGATGTGGCCACTCAAGAATTGGCCAGACAAGGGGAGGAAATCCTCTCTCAGCTATAC...CGA--     137
B_EHOA        AGGGGGTTGCCAACACTCAAGGATTGGACAACCGGGGGGAGGAAATCCTCTCTCAGCTATAC...CGC--     137
SD_MM251      AGGCGGATGCAACCACTCCAGAATCGGCCAACCTGGGGGAGGAAATCCTCTCTCAACTATAC...CGC--     137
STM_STM       AGGCGGTTGCCGCCACTCCAGGATTGGCCAACCAGGAGGAGGAAATCCTTTGGCAACTATAC...CGC--     137
VER_AGM3      ..........................................GACTTGATTGAACAACTCAAA...GCA--      65
GRI_AGM677    ....................................................CAGGACTTGCAT...AGG--      50
SAB_SAB1C     ..........................................................GAGCTGCAT...CGG--      65
SYK_SYK       GAACATTCCTAGAGAAA.....................................................GGA...--      80

TAT_CONSENSUS ?c?aac?cc?TGtaataa?a??TGctattGtAAaa??TGttgcTatCAtTGcca??ttTGcTTc?taaaa      95
A_U455        TAC---TG-T----GC--C...--T---------GTG-------GG--------AC-------TC-G---     120
B_HXB2R       TAA---TG-T----CC--T...-------------AG-------T---------AG----T---A---C-     120
C_UG268A2     TAA---T--T----C-...-AG--T-T-G------AA---A-T--------T--AG-------TC-G--C      80
D_ELI         TAGG--T--T-----C...-AG--TC---------AG----------------CAG--------T----C     120
F_BZ163A      TAC---T--T----CA...-AA------------CGA-------T------T--ATGG-----TAC--CG     120
O_ANT70C      C-AG-TC--T--------T...--------C----GA--C-----------T-TG----T---G---G-     120
O_MVP5180     C-A---C--T--------T...--------C----GA--C-----------T-TG----T---AC---G     120
CPZGAB        AAG---AGTT--------T...--------C---GCT-----T-----C--TATATA-------AC----     120
CPZANT        AGC---G--A--C-----C...--T--C------TGC-----T--C-------CTC-C------AC----     120
A_ROD         C-T-GAAA-A--C-----CTCA----------GCGA--C-----C-----T--GA-G--T--TC----C     207
B_EHOA        T-TT-AGG-A--C-G---T-CA-----C-------AA----C---C--------GC-------TC-T---     207
SD_MM251      T-TTGAGG-G--CT----C-CA-------------AG--------C--------GT----T--TC-T---     207
STM_STM       A-T-GAGGAG---TG---C-AG----T---C----AA----------C-----GC-G--T--TG-T-C-     207
VER_AGM3      C-TG-AG-GG----CA--C-AG------------TGT--C--T-----C--T--GC-------TT--C--     135
GRI_AGM677    ATT-CAA--A----C---T-AA--------C----AA--C--T-----C--TG-GC--------C-GC-G     120
SAB_SAB1C     G-TCCAGG-C---G----T-CT----TC--C---GTT--C----T------TATTC-C------CAT---     135
SYK_SYK       A-C---A--A------...-AA----T---C----AC----------------GC----T---C-TC--     147
```

```
                  \/
                  \/     \/    Rev ->    \/
TAT_CONSENSUS AAgGgcttaGG?aT?tcctATggcaggaaga???g???????aga????g?cgaaga????ctcct?aaa   143
A_U455        --A--------C--T----------A-----AGC-G......-A-CCCC-A---G--...C-----C--G   181
B_HXB2R       --A-C------C--C----------------AGC-G.....---CAGC-A------...G---A-C-G-   181
C_UG268A2     --A--------C--C----------------AGC-G.....---CAGC-A----AC...G-----CC--   141
D_ELI         --A--------C--C----------------AGC-G.....---CAGC-A---G--...C-----C--G   181
F_BZ163A      ----------C--C----------------AGC-G......-A-CAGC-G-ACC--...A-----C-G-   181
O_ANT70C      -----T--G--A--C----------------AGC-A......G--.....---CC-...G--G--GC--   175
O_MVP5180     -----T--G--A--C---C------------AGC-A.....---...A-A-C-GC-...G--G--GC--   178
CPZGAB        --A--------A--C----------------AGC-A......-CCACAA-A---C-CACGG-----GC-G   184
CPZANT        --A-----G--G--T----------------GAA-AGCTCGA--GAACA-A----C-...A--G--G---   187
A_ROD         -----GC-C--G--A-GT----AAC-A---GGCA-A......C--.....A----G...A----AA-G-   262
B_EHOA        ------C-G--G--A-GT----AAC-CTC-.............---.....AAG---...T--T-AA---   256
SD_MM251      -----A--G--G--A-GT----AGCA-TCA.............C--AAGA-AA------...A----GA---   262
STM_STM       --A--TC-T--A--AA------AGC-ATCA.............---.....A------...GT-AAGA---   256
VER_AGM3      -----A-----TG-TA-----CATGCC...CCTA-G......-TC.....A------...AAGAAGATTG   187
GRI_AGM677    ----------AG-GCGT---CATGTCTCT.............--G.....AA----...AGAAAAAC--   169
SAB_SAB1C     ----C-C----C--TCG----TATGTC...CCTA-G......CC-.....A------...G--T--A-G-   187
SYK_SYK       -----A--G--A--CA-T----C---A...CCAA-G......-AG.....---GC-...G---GCAGT-   199

                                                  exon \/ exon
TAT_CONSENSUS gca?t?aggat?atcaag?tcctgtAcca???aagcaaccc?tatccaaa????g?ggg?ac...a??ca   194
A_U455        ---G-A-----C-----AC--T-A-----...---------T-GC--C--TCCCAGA--GT-...TCGGC   245
B_HXB2R       A--G-C--AC-C------C-T--C--T--...---------ACC---C--CCCC-A---G--...CCGAC   245
C_UG268A2     ---G-G-----C-----AA--T-A--T--...---------T--C--C--CCCC-A---G--...CCGAC   205
D_ELI         --GG-C---C-C------T----A-----...---------TCC---C-GCCCC-A---G--...CCGAC   245
F_BZ163A      ---GCC--ATAC------A--T-------...---------A-----C--GCCC-A---A--...CCGAC   245
O_ANT70C      --CA-CCA---C--A---A----------...------T--CC-A---TC...ACCAA-AGG...-AG--   236
O_MVP5180     --TA-CCA---A--A---A----------...G-----T--C-G---C-C...ACC--CAGG...-AG--   239
CPZGAB        ---G-A-AA--A-----A-T--A-----...---------T-----C-GTCCA-A---A--...-AGG-   248
CPZANT        ---G-G-AA--A------A-------T--...--ATCC-TACCCAAA-CCAGCA-GATCC-G...-GCAG   251
A_ROD         AA-C-A--AC-C---CGTC----AC----...G-CA--T--A------C-AGGACC---G--...-GC--   326
B_EHOA        -GGC-A--AC-AC-AC-TC-T---C----AAC--CAG-T--C-----GC-AGGACC---G--...-GC--   323
SD_MM251      AGGC-A---C-A--AC-TC-T---C-T--...--CA---T-A--C----C...A-GACCCGG...CACTG   323
STM_STM       AGGC-A--AC-T---C-AT-T---C-T--AAC--CAG-T--C------C-AGGGCCA--A--...-GC--   323
VER_AGM3      CTCCGCTT---...-GCTT-----A--A-...--A---T--A------CCCGTG-AA--G--...-GC--   248
GRI_AGM677    ---C-C-A---A------AC---A---GG...C-----T--A------C-GTTCAGC--A--...GGC--   233
SAB_SAB1C     AG-T-TC-C--A------T-T--C---AC...--C                                     219
SYK_SYK       TTTC-G-----......-A-AGC-C---T...-CAGG-A--C--C---GG...GCC--AAGG...-CC--

TAT_CONSENSUS ?g?accg?ag?Aa?agaaGaagaaG???GTGGaga??aaggcaga?aCAgatcc??????????????   237
A_U455        A-GC-A-G-AG--TC----------...--------GC------A-G-------GA.............   297
B_HXB2R       A-GC---A--G--T----------...--------GAG--A----G--------A.............   297
C_UG268A2     A-GC---G--A--TC----------...--------GC---A----G--------G.............   257
D_ELI         A-GC---A--G--C----------...--------GCG-------G--------G.............   297
F_BZ163A      A-GC---A--A--TC-------G--...--------GC--------G--------G.............   297
O_ANT70C      G-A--GCC--G--G-AC--G--G--GAG-------AG-------GCC---G-GGATACCCTCGCCGCAAG   306
O_MVP5180     GAA--GCC--G--G-AC--G-----AAG-------AGG--A---GCC--AG-GGACAGCCTTGCCACCAG   309
CPZGAB        A-GCT--G--A--TC--C----G--...----C--GC---A----AG------AG.............   300
CPZANT        CCAGAA-A-AT--G-----...---...-----AGAA---A---GAT---G-GGA.............   300
A_ROD         GCC-A--A--A--C---------C-...-----AGCA-C--TG--G------A-T.............   378
B_EHOA        GCC-A--A--A--C--------G--...-----A-CA-C-AGG-CT-------TTGGCCCTGGCAGATCG   390
SD_MM251      CCA---AG--A-GGCA------G--ACG-------AGGC--TG-CA----C---T.............   378
STM_STM       ACC-AA-A--G--C--------G--...-------CAG---TG--AT--AC---T.............   375
VER_AGM3      GAC-A--C--A-GGGCC--G-----...-------CA-GC---AGA----C---GAGCCTTGGCAGAAAG   315
GRI_AGM677    GAC-A--G--G-GGGC----CAG--...------                                      264
SYK_SYK       G-C-AACCC-C-GAC----------GCA-------CAG--AAG--CT--ACGT-TCTACCT.........   315

TAT_CONSENSUS ??????????????????????????????tg????ga?t?a                            243
A_U455        ............................-TC......GCT                               303
B_HXB2R       ............................-TC...--T-AG                               306
C_UG268A2     ............................--A...--T-A-                               266
D_ELI         ............................--AGATTGC-G-                               309
F_BZ163A      ............................--AGATT-G-AG                               309
O_ANT70C      GGTTCT...TGCCACTGTTGTACACGGACC-CA...--ACA-                               342
O_MVP5180     GATTCTTGCAACAGTTGTACACGGATCTCAG-A...C-A-A-                               348
CPZGAB        ............................--A...--T-AG                               309
CPZANT        ............................G-G...CCT-GC                               309
A_ROD         ............................G-CCCT-GCCG-                               390
B_EHOA        AAT............ACATCCACTTCCCGA-TCGCCA-C-G-                               420
SD_MM251      ............................G-CCTT-GCAG-                               390
STM_STM       ............................G-CCTT-GCAA-                               387
VER_AGM3      AATCTGGCACAGCAGAGTGGAAGAGCAACTG-T...-CAAGC                               354
SYK_SYK       ......................TCAGCAGAG...A-TCTT                               330
```

HIV1 TAT CONSENSUS

```
                                <- vpR
CONSENSUS-A   ATGGA?CC?GTAGATCCTAaCCTAGAGCCCTGGAAcCA?CCGGGAAGTcAGCCTACAACTgCTTGTAgCA      67
CONSENSUS-B   -----g--A----------ga-------------g--T--A------------ag----------c--      70
CONSENSUS-C   ???-?-??-?---------??-------------------T--A------------A----------?T-      58
CONSENSUS-D   -----t--a-----------------------------T--A------------gg---c------A--      70
CONSENSUS-F   -----G-TA-----------T----T--------?--T--A-----C------------C------CA-      69
CONSENSUS-O   -----T--A----------G-G?-GCCC--T---C-?--C--T--?--??----CCA?-?CC------AT-     63
CONSENSUS-U   -----T--A-----C-----A-----------------T---------------A----------C--      70
CONSENSUS-CPZ ---------?----??--G-???????--?---??A--C--??????-???--A??---????--?-AT-     39

                                                                    3'sj  3'sj     rev cds
                                                                     \/    \/       ->
CONSENSUS-A   AgTGTTAcTGTAAAatgTGtTGCTggCATTGccaA?ttTGCTTTCTgaacAAaGGCTTAGGCATcTCCTA    136
CONSENSUS-B   -t--C--T-------a--------tt---------G----t--caca-cA--------------------    140
CONSENSUS-C   -------T-------aa---A---At-----t-t-G---------a--ca--------------t-----    128
CONSENSUS-D   -----c-T-------A--------AT---------g-------Ca-A-cg--------------------    140
CONSENSUS-F   -A--C--T------CGA-------TT-----T---TGG------ACA-C?--G----------------    138
CONSENSUS-O   -T--C--T--C----GA--C----AT------T-TG----T--C??A-??--G--T--G--A------?-    128
CONSENSUS-U   -------T-------A--------AT-------C-G-G--------A----------------------    140
CONSENSUS-CPZ -?--?--?--?---???-----T-A?--?--????-??-----CACA--A--------?--?--?-----     90

              \/ 3'sj
CONSENSUS-A   TGGCAgGAAGAAGCGG......AgaC?cCGACGAgGAaCTCCTCAaaGcagTaAGGAtCaTCAAAaTCcT    199
CONSENSUS-B   ----------------.....----AG------a--g--------ga----c--ac-------gt-t--    204
CONSENSUS-C   ----------------.....----AG------A-CG------c-------G---------------?-    191
CONSENSUS-D   ----------------.....----AG------a--c--------g--g--C---C-------G?----    203
CONSENSUS-F   ----------------.....-A--AG--?-ACC----------G-----?C--ATA------G---T-    200
CONSENSUS-O   ----------A.....?-??---?-?C-G--G--GC----?A-CCA---?--A--G-----    181
CONSENSUS-U   ----------.....----C-------A--T---------A---CG----------------    204
CONSENSUS-CPZ ----------???-A?C????-??A-??----?AC??--?--G?-?-----?-AA--A-----G--?--    139

              exon \/ exon
CONSENSUS-A   aTAcCAAAGCAAcCCtTaCCCcaaaCccaagGGgtC???cCGaCAGgCCcGaAAGAATCGaAGAAGaAGG    266
CONSENSUS-B   c--t----------?Cct----gc---g-----A-...----------G----------------    270
CONSENSUS-C   ---T----------g-----A-...----------G-g----------------    258
CONSENSUS-D   ----------CCT----gc---G-----A----------G----A----------    270
CONSENSUS-F   G----------A--T-----G---G----AA-...----------G?----------G---    266
CONSENSUS-O   G-----?-----T--C???--??C---???A--AAG...-A-?A-C---A-G-----CA-G--?--?---    236
CONSENSUS-U   ----------T--C-------C-----GA-----...T----------G----------G---    271
CONSENSUS-CPZ ?--?----?????--?--?--?-??--AG??--A?-...?A-????--??-?-GA---????----?---    178

CONSENSUS-A   TGGagAGCAAGaCAgAGACAGATCgattCgcTTAG                                    301
CONSENSUS-B   -------aG----------Cgg---a----                                        305
CONSENSUS-C   ----------?--?------c-----A----                                       291
CONSENSUS-D   ----------G----------c-----A--g-TGA                                   308
CONSENSUS-F   ----------G--??--?-----CG-G?-?----T??                                 297
CONSENSUS-O   ------AG?--?---GCC--?G-G--?A?C--?GCC?C?AGG?TTCTTGC?AC?GTTGTACACGGA?CTC  294
CONSENSUS-U   ----------G----------A----                                            306
CONSENSUS-CPZ ---?????-------?A?---?-????G???--??                                  196
```

HMMER Sequences in the Rev Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_HXB2R	HIVHXB2R	K03455	Starcich,B.	Science 227, 484 (1985)
D_HIVELI	HIVELI	K03454	Alizon,M.	Cell 46, 63 (1986)
F_BZ163A	HIV1BZ163A	L22085	Louwagie,J.J.	ARHR 10, 561 (1994)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68, 1586 (1994)
O_MVP5180	HIVMVP5180	L20571	Gurtler,L.G.	JVI 68, 1581 (1994)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
CPZANT	SIVCPZANT	U42720	Vanden Haesevelde,M.	Virology 221, 346 (1996)
A_ROD	HIV2ROD	M15390	Clavel,F.	Nature 324, 691 (1986)
B_EHOA	HIV2EHOA	U27200	Rey-Cuille,M.A.	Virology 202, 471 (1994)
SD_MM251	SIVMM251	M19499	Franchini,G.	Nature 328, 539 (1994)
STM_STM	SIVSTM	M83293	Novembre,F.J.	Virology 186, 783 (1992)
VER_AGM3	SIVAGM3	M30931	Baier,M.	Virology 176, 216 (1990)
GRI_AGM677	SIVAGM677	M66437	Fomsgaard,A.	Virology 182, 397 (1991)
SAB_SAB1C	SIVSAB1C	U04005	Jin,M.J.	EMBO J. 13, 2935 (1994)
SYK_SYK	SIVSYK	L06042	Hirsch,V.M.	JVI 67, 1517 (1993)

The following alignment was generated using the HMMER program as described in the introduction to this Part and in Part III. For simplicity, only representative types and subtypes are shown. An ordinary consensus sequence (lowercase signifies majority, uppercase signifies 50% or greater) was created from these sequences using MASE; this is not a "most likely sequence" generated from the HMM model (Part II). Annotation is based on HIV-1s, therefore users should be cautious about its applicability to other PIV sequences.

```
                    \/3'sj
REV_CONSENSUS ATGgcaggaagaagc?????????gaa???ga?gaaga?ct?ct?aaagca?t?ag??t?atcaag?tcc     48
A_U455        ------A--------GGA......A-CCCC--C--G--C--C--C--C--G---G-A--GA-C-----AC--T     64
B_HXB2R       ---------------GGA......--CAGC--C-----G--CA-C-G-A--G-C--AC-C------C-T-     64
D_ELI         ---------------GGA......--CAGC--C--G--C--C--C--G--GG-C--GC-C------T---     64
F_BZ163A      ---------------GGA......A-CAGC-GCACC--A--C--C-G----GCC--ATAC------A--T     64
O_ANT70C      ------------............AGCGAG--C--CC-G--G--GC----CA-CCAGA-C--A---A---     58
O_MVP5180     ---------------.........---GAA--CC-GC-G--G--GC----TA-CCAGA-A--A---A---     61
CPZGAB        ---------------GAACCA...C--GAC--C-C-CGG--C--GC-G---G-A-AAA-A------A-T-     67
CPZANT        ------------GAAGAGCTCGAA-G-ACA--C---C-A--G--G------G-G-AAA-A------A---     70
A_ROD         ---AAC-A---G............-C-GAC--A----GA--C-AA-G-AA-C-A--AC-C---CGTC---     58
B_EHOA        ---AAC-CTC--.....................--AAG---T--T-AA----GGC-A--AC-AC-AC-TC-T-     52
SD_MM251      ---.......--C--TCAC......---AGA--A-----A--C-GA---AGGC-A--GC-A--AC-TC-T-     58
STM_STM       ---.........---.........--TCAA--A-----GT-AAGA---AGGC-A--AC-T---C-AT-T-     52
VER_AGM3      ---.........CC-.........CT-GGATCA-----A...AGA-G-TTGC-CC-CT-G---GCTT---     49
GRI_AGM677    ---.........TCT.........CT-GGAA-A-----A...AAAC-----C-C-AGA-A------AC--     49
SAB_SAB1C     ---.........TC-.........CT-GGCC-A-----G--T--A-G-AG-T-TC-CA-A------T-T-     52
SYK_SYK       ---.........CCA.........--CCAA-GAAGC--G...-AGCTC---G-ATTTC-G-GG-T-A-AG     49

                                        exon \/ exon
REV_CONSENSUS tgtAccaaagc?????????????????aAtcC?tatCCaaa?cc?g?aGGaac??g?aa?gcccGaAg     92
A_U455        -A---------.....................--C--T-GC--C--T--CAG---GT-TC-GC-G---A-G-A    116
B_HXB2R       -C--T------.....................--C--ACC---C--C--C-AG--G--CC-AC-G--------    116
D_ELI         -A---------.....................--C--TCC---C-GC--C-AG--G--CC-AC-G--------    116
F_BZ163A      ----------.....................--C--A-----C--G--C-AG-----CC-AC-G--------    116
O_ANT70C      ----------.....................-----CC-A---TCA--AAG-----GCA-G--C---A-G-A    110
O_MVP5180     -------G---.....................-----C-G---C-CA--G-C-----GCA-A--C---A-G-A    113
CPZGAB        -A---------.....................--C--T-----C-GT--A-AG-----AA-G--G--T--G--    119
CPZANT        ----T---.....................-----C--C-----A--A-C----T-CA-AGCA---A----    119
A_ROD         -AC----G-CA.....................-----A------C-AGGACCG--G--AGCC-GCCAA-----    110
B_EHOA        --C------CGAGTGAGTATGGCACAACAG----C-----GC-AGGACCG--G--AGCC-GCCAA-----    122
SD_MM251      --C-T----CA.....................--CT-A--C----CAGGACCC--C--TGCC--CCAGA----    110
STM_STM       --C-T----CAACA..............G----C------C-AGGGCC------AGCC--CCAAA----    107
VER_AGM3      --A--A---A-.....................-----A------CCCGTG-A---G--AGCC-GACAA--C--    101
GRI_AGM677    -A---GGC---.....................-----A------C-GTTCAGC-----GGCC-GACAA--G--    101
SAB_SAB1C     -C---AC--CT.....................-----A------CCTGGACA---G--AGCC-GACAAA----    104
SYK_SYK       C-C---T-CAG.....................G-A--C--C---GGG--G-A----C-CA-GC-AA----C--    101

REV_CONSENSUS ?aa?agaaGaag?agGTGG???aga?a?aggcaga?acagatcc??????tt???gca?ggat?cTt??c    140
A_U455        G--TC-------A------...---GCA-----A-G-------GATTCGC--AGT-AGC----T---AG-    183
B_HXB2R       G--T--------A------...---G-G--A----G--------ATTCGA--AGT-A-C----C---GG-    183
D_ELI         G--C--------A------...---GCG-------G--------GTGAGA--GCT-A-C----T--CGG-    183
F_BZ163A      A--TC-------G------...---GCA-------G--------GTGAGA--AGTAG-C----T---AG-    183
O_ANT70C      G--C--G--G--G------...---AGA------GCC---G-GGATACCC-CGCC---A--G-T---GC-    177
O_MVP5180     G--C--G-----A------...---AGG--A---GCC--AG-GGACAGCC--GCCA-CA----T---GCA    180
CPZGAB        A--TC--C----G------...C--GCA--A----AG------AGTGAGA--AGT-G-A--G-G---GC-    186
CPZANT        A--T-AG-----A------...-AGA-A--A---GAT---G-GGAGGGCC--GCA---C--G-T---CG-    186
A_ROD         A--C--------AC-----...-AGC-AC--TG--G------A-TG...GCCCTG--CGAT-GCA-ATAT    174
B_EHOA        A--C--------G------...-A-C-AC-AGG-CT-------TTG...GCCCTG---GATCGAA-ACAT    186
SD_MM251      GC-A-----G--AC-----...---AGGC--TG-CA----C---TG...GCCTTG---GAT-GAA-ATAT    174
STM_STM       G--C--------G------...---C-G---TG--AT--AC---TG...GCCTTG---AAT-GAA-ATA-    171
VER_AGM3      A-GGGCC--G--A------...---C-AGC---AGA----C---GA...GCCTTG---GAA-GAA-CTGG    165
GRI_AGM677    G-GGGC----CAG------...---A-GCA----CAG--A--TGATAAGA--GCA-G-A-AG-C--CAA-    168
SAB_SAB1C     A-GAGC---GCAGC-----GCT-A-C-GC-A---CAGGTC--A-AC...T-GGCA-AGA-A--T---...    168
SYK_SYK       ACGA--------GCA----...---C-G--AAG--CT--ACGT-TC...TACCTTCAGCA--GAA-C...    162
```

```
REV_CONSENSUS ac?ttt???????ga????cgga?cct???cc?????t??a??t?ccacc?cttga???a?tt????t??   173
A_U455        GAT-GCCTGGGAC--CCTG----A---GTG--T...C-TC-GC-A-----GA----GAG-C--CGCC-TG   250
B_HXB2R       --T-A-CTGGGAC--TCTG----G---GTG--T...C-TC-GC-A-----G-----GAG-C--ACTC-TG   250
D_ELI         --T-A-CTGGGAC--CCTG----G---GTG--T...C-TC-GC-A-----G-----GAG-C--AATC-TA   250
F_BZ163A      T-T-G-CTGGGAC--CCTGA---A---GTG--T...C-TC-GC-A-----A-C---GAG-C--ACTA-TA   250
O_ANT70C      --TG--GTACACG--CCTCA-A-CAA-AATATT...G-GG-CT-A-----T-----GCA-C--AGCA-CA   244
O_MVP5180     --AG--GTACACG--TCTCA---CAA-AAT-TT...G-GG-CT-A-----T-----GCA-C--AATA-CA   247
CPZGAB        --T-A-CTGGGAC--CCTC--A-A---GGGGAT...T-GG-GC-A---GAG-----CAAGC--AGCC-GC   253
CPZANT        --TC--                                                                   192
A_ROD         --A---............C-T--T--G...--AGCT......GATT-G--T-----CCAGAC-...A-AC   220
B_EHOA        C-AC--............C-C--TT-G...--AACT......GAGGG---T-----TTTGGC-...A-AC   232
SD_MM251      T-A---............C-T--T--G...--AACT......GATA-G--T-----CTTGGC-...A-TC   220
STM_STM       T-A---............C-T--T--G...--AGCT......AGTA-G--T-----CTTGGC-...G-AC   217
VER_AGM3      CACAGC......A--GTGGAA--G-AA...-TG...G-GC-AGCG......A----CCA-T-GGTGC-CG   217
GRI_AGM677    --C--C............GA---T-AACAA-TGGTGGCTC-AC-T-A-GAG--GC-GCTTGAGAATAAGG   226
SAB_SAB1C     ................GA-ACA---GTGT-A...CAGATTGAT...-ACT-G-CTCAGGAATTTGACC   214
SYK_SYK       ................TTT--GG-A...ATC...T-TGGAAGC......AGGACAGC-GC-...T-GG   199

REV_CONSENSUS a??a??????g??????????a??atct??ct????ag??????c?g??????????????????????cc   189
A_U455        -TTGTAGCGA-.........AGCTG-GGAA--TCTGG-ACG...-A-CAGCCTCAAGGGACTGAGACTGG   308
B_HXB2R       -TTGTAACGA-.........G-TTG-GGAA--TCTGG-ACG...-A-.....................GG   287
D_ELI         -TTGCAGTGA-.........G-TTG-AGAA--TCTGG-ACG...-A-.....................GG   287
F_BZ163A      -TA-CAGCGA-.........A-TTG-GAACAGGGGGCTGAG...GA-.....................GG   287
O_ANT70C      GGG-TCCAGAA.........GGTG---AGCTATCTG--GCT...TG-.....................A-   281
O_MVP5180     GGG-TCCGGA-.........GCTG---GACTACCTGG-ACT...GG-.....................A-   284
CPZGAB        -ATGTGTGGA-ACAACTCAAG-CGT-GGGA-ATCTA-TACTTCA-A-.....................--   302
A_ROD         -GC-T...CT-.........C-GGGA--TA--ATCC--GAG...-TT.....................--   254
B_EHOA        -GCGG...CT-.........C-GA-C--TAT-ATCA-AGAC...-TT.....................--   266
SD_MM251      -GC-A...CT-.........C-GA-C--TG--ATCG--AGC...ATA.....................--   254
STM_STM       -AC-A...CTT.........C-GGGCT-GT--ATCC--GAT...-TA.....................--   251
VER_AGM3      -TC-G...............C-GC----GG--ATAC-ACAG...TT-.....................--   248
GRI_AGM677    -CTTGGTGTTA.........C-AC----G.........................................--   248
SAB_SAB1C     -GTTGGTTCTC.........G-CA----T......C--CAA...-CT.....................--   245
SYK_SYK       -GG-CAGTCT-.........C-GC-G--TCAGATCTCTGAT                               231

REV_CONSENSUS ????ga?cctcc??????????????????????c??a?t?ct??aga??????????????????????   204
A_U455        GGTG-GAGGG--T.....................-AA-TAT--GTG--A.....................   336
B_HXB2R       GGTG-GAAGC--T.....................-AA-TATTGGTG--A.....................   315
D_ELI         GGTG-GA-A---T.....................-AA-TAT--GTG--A.....................   315
F_BZ163A      GGTG-GAAGC--T.....................-AA-C-T--GGG--A.....................   315
O_ANT70C      TGTG--T---AGGGCAGAAGATAATTAATGTTTG-AG-A-TTGTGCAGC.....................   330
O_MVP5180     TGTG--T---GGG.....................ACA-AAGACAAT                         309
CPZGAB        T.........-AG.....................ACTGC-A-GGG---GACTGTGCCTGCT.........   333
A_ROD         T...--C-----A.....................ACTCA-CTACC---A.....................   279
B_EHOA        A...A-C-----A.....................ACC-G-A--CC-ACCGCTCAGGCTTCCACCTGCATA   312
SD_MM251      A...--T-----A.....................ACC-A-A--CC---GGCTCTC...............   285
STM_STM       A...--C-----A.....................-CC-ACCT-CC-A-G.....................   276
VER_AGM3      T...--C-----T.....................AGTTCAT--                             267
GRI_AGM677    T...--C-----T.....................-AT-T-CA-                             267
SAB_SAB1C     ATCGCTG-----A.....................GGACA-C-AAC---G.....................   273

REV_CONSENSUS ????????????????????????????????????????????????????????????????????   204
A_U455        ........................................................................   336
B_HXB2R       ........................................................................   315
D_ELI         ........................................................................   315
F_BZ163A      ........................................................................   315
O_ANT70C      ........................................................................   330
CPZGAB        ........................................................................   333
A_ROD         ........................................................................   279
B_EHOA        CCTCCGATATGGGATCAGCTGGTTCCAAGAAGCAATCCAAGCAGCAGCCAGGGCTGCGGGAGAGACTCTT   382
SD_MM251      ........................................................................   285
STM_STM       ........................................................................   276
SAB_SAB1C     ........................................................................   273

REV_CONSENSUS ??????????????????????????t?tcat?c??ta?tgg?g?cagga??????????????????   221
A_U455        ........................-C-TC-G-TG--T---G-T-G---.....................   360
B_HXB2R       ........................-C--C-A-AG--T---A-T-----.....................   339
D_ELI         ........................-C--C-A-AG--T---A-T-----.....................   339
F_BZ163A      ........................-C---CG-TG--T---G-T-----.....................   339
O_ANT70C      ......................-G-A-CA-AA--C---CTA--A--.....................   354
CPZGAB        ....................GGGGGGAA-T--T-AA--T---G-AA----.....................   363
A_ROD         ........................-C---G...AG-C--.............................   291
B_EHOA        GCGAGCGCGGCGAGGACCTCGTGGGGAG--C-.....CA--A-AGC--CAGGAGAGATCATTGCAATAC   446
SD_MM251      ......................TGCGAC-C-A-GAAGGGTTC-AG-A-T.....................   312
STM_STM       ........................G--.....C-CCA-GATACT.........................   291
SAB_SAB1C     ......................AAC--GA-AGCTAACAGCT--TCT                         297
```

HIV1 REV

```
REV_CONSENSUS ????????????actaa???????ga?                    228
A_U455        .............-----A......A-T                   369
B_HXB2R       .............-----A.......-A                   348
D_ELI         .............---G-GGAACAGTGC                   354
F_BZ163A      .............-----A.......-A                   348
O_ANT70C      .............-T-GCA.......-A                   363
CPZGAB        .............G----A......A-T                   372
A_ROD         .............G--G-G......ACT                   300
B_EHOA        CCAGGAGGATCAG--A-GG......-GC                   468
SD_MM251      .............C--C-G.......-C                   321
STM_STM       .............G-AG-G......A-T                   300
```

```
       tat ->       \/ 3' sj
CONSENSUS-A   ATGGCAgGAAGAAGCGGAgaCagCGACGAG...GAaCTCCTC...AAaGCAaTAAGGATCATCAAAaTCc     64
CONSENSUS-B   ----------------------------a...--g------...--ga--G-c--ac-------gt-t-     64
CONSENSUS-D   ----------------------------a...--C------...--G--gg-C---C-------G----     64
CONSENSUS-F   ------------------A------?-ACC...---------...-G----G?C--ATA------G---T     62
CONSENSUS-O   ----------------A?--?...---.....C-G-?G--G???C----?--CCA---?--A--G----     53
CONSENSUS-U   --------------------GC-------...---------...-G--TCG-C---------------     64
CONSENSUS-CPZ ------------???-A?C???AA-?????G?????-??--G????-?---G-?-AA--A-----G--?-     44

                    exon \/ exon
CONSENSUS-A   TATACCAAAGCAAcCCaTaCCCCAAaCCCAagGGg...TCtCGaCAGGCCcGaAAGAATCGaAGAAGAAG    131
CONSENSUS-B   -c--t-----------?cct----gc---g-----...A-C-------------G---------------    130
CONSENSUS-D   ----------------TCCT----gc---G-----???A-C-------------G----A----------    131
CONSENSUS-F   -G-----------------T-----G---G----A...A-C-------------G?-----------G--    128
CONSENSUS-O   -G-----?-----T--C???--??C---???A--A...AGCA-?A-C---A-G-----CA-G--?--?--    108
CONSENSUS-U   -G--------------T-----ACC---AG-----...A-C------A------------------G--    131
CONSENSUS-CPZ -?--?---???--?--?--?--?-??--AG??--A...?-?A-????--??-?-GA---????----?--     84

CONSENSUS-A   GTGGAGAGCAAGgCAgAGACAGATCGATTCGcTTAGTgAGCGGATTCTTAGCacTTGCCTGGGACGACCT    201
CONSENSUS-B   --------ag--a------------Cgg---a-------at---------------aT---------t--    200
CONSENSUS-D   ---------g---------------c-----A--G----A--------c-------aT-----------    201
CONSENSUS-F   ---------------??--?-----CG-G?-?-----??A------------T----T-----------    191
CONSENSUS-O   -------AG?--?---GCC--?G-G--?A?C--?GCC?C?A--?-----GC?--?GTTG-ACACG--?--    166
CONSENSUS-U   ----C-------A-----------AGGG--A-----------T----------------------T--    201
CONSENSUS-CPZ ----?????---A---?A?---?-????G???--???-?A?--G-?---??----??T?----------    128

CONSENSUS-A   GCGGAgCCTGTGCCTCTTCAGCTACCACCGcTTGAGAGACTTCatCTTGATTGtaGCGAGgaCtGTGgAA    271
CONSENSUS-B   ----------------------------------------Ac----------a------t-------    270
CONSENSUS-D   -a---a-----------------------------------A-----A----C--------T-------    271
CONSENSUS-F   -A---A-------------------?--A-?----------A?-A--A--A?C------?-T----A-.    255
CONSENSUS-O   CA-?-CAA-AAT?T-G-GG-CT-------T------CA----A??A-CAGGGA-CC?--?-?TGA-C??C    227
CONSENSUS-U   ------------------------------A----------TG---------C--T----G--------    271
CONSENSUS-CPZ C--A-A----G-GA-T-GG--------GA------C-AG---AGC--GC-A---GTG---ACAAC-CA-G    198

CONSENSUS-A   CTTCTGggACgCAGcag?ctca?ggg?ctgagactGGGGTGGGAaGGccTCAAgTAtCTGTGGAAtCTtC    338
CONSENSUS-B   --------------.................???----------c------a---tg---------C-    319
CONSENSUS-D   --------------.....................----------c------A----------C-    320
CONSENSUS-F   ..CAG---G-T?--G--.................----------C------ACT----G-------CA    304
CONSENSUS-O   TA?---?-?-T?G-ACTGTGG-TCCTAGG-CAGAA-ATAATTA-T-TTTG--GAAT-TGTGCAGC-G-AA    293
CONSENSUS-U   ----------A------C----A---A------------------C----C-A------G---------    341
CONSENSUS-CPZ ACGT------ATCTA-TA--TCACA-C--C------CTAC----GA-A--GTGCCTG---G--GGAA--A    268

CONSENSUS-A   TGcTgTaTTGGGGTCGGGAACTaAAAAt?G                                          367
CONSENSUS-B   ---a-------a---a---------G-a-AG                                        349
CONSENSUS-D   -a-A-------a---a------G-gG-A-AGTGC                                     353
CONSENSUS-F   C--------------A---------G-ATAG                                        335
CONSENSUS-O   CA-??-????-?T?-??-?-????-G-?-ATTGA                                    310
CONSENSUS-U   -A--A--C-------A---------G-ATAG                                       372
CONSENSUS-CPZ -T-AA--------AAA---G--------                                          296
```

HMMER Sequences in the Vpu Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_SF2	HIVSF2	K02007	Sanchez-Pescador,R.	Science 227, 484 (1985)
D_ELI	HIVELI	K03454	Alizon,M.	Cell 46, 63 (1986)
F_BZ163A	HIV1BZ163A	L22085	Louwagie,J.J.	ARHR 10, 561 (1994)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68, 1586 (1994)
O_MVP5180	HIVMVP5180	L20571	Gurtler,L.G.	JVI 68, 1581 (1994)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
CPZANT	SIVCPZANT	U42720	Vanden Haesevelde,M.	Virology 221, 346 (1996)

The following alignment was generated using the HMMER program as described in the introduction to this Part and in Part III. For simplicity, only representative HIV-1 subtype sequences are shown. An ordinary consensus sequence (lowercase signifies mqjority, uppercase signifies 50% or greater) was created from these using MASE; this is not a "most likely sequence" based on an HMM model.

```
VPU_CONSENSUS ATGc?tcattt?ga??t?ttagcaaTaa?agcattaatagtagtactaataata?caaTagttgT?TGGa     62
A_U455        ---ACA-C---G--AA-C-GG-------C--GGC-G-------C------CT--G-------A--G----     70
B_SF2         ----AATC---AC-AA-A---------GT-T-----G--------GC-------G----------G----     70
D_ELI         ----AA-C---A-GGA-AA--------GC-------G------C-A-----C--G----------G----     70
F_BZ163A      ---TC-G----GTTAGCAA--AGTG---C--------------C----------G----------G----     70
O_ANT70C      ----A----AGG--CC-GC---------T-ATTA-T-GT-CTT-G--GT-T---AATG--A--T-A---G     70
O_MVP5180     ----A---AGAGA-CC-GC----CT---T---T----GT-CTT-GTGTC-T---AATG--C--A-A---T     70
CPZGAB        ---AC--TG--A-TCGGC----TTC-C...AT----G---GCT--A-TGC-TGGAAC--TTG-A-A---G     67
CPZANT        ---AC-A--A-ATTTGAG-AT--TT-T..................--TGCCT-TAGT-----AC-A----     52

VPU_CONSENSUS c?atagta?ttata?????????gaatatagaaaaa?ataaaga?????????cA?aga?aAatagatag    109
A_U455        -T------GG----.........-----A-A-----AT-GCTA-AGC......A-A-A-A--TAGACAG-    125
B_SF2         -C------C-C---.........--------G----T--T----.........--A---%-------C--    121
D_ELI         -C------T-C---.........----------GG-T-A---AG.........--A--GAG------CT-    122
F_BZ163A      -T------TA----.........--------G---CTG-T---G.........--A---A-----A----    122
O_ANT70C      GGT-TA-TC---G-AAATATTTA---C-A-AGG--CA-G-C---AAG......G-G---G-----CT-GA    134
O_MVP5180     TGT-TAACC---G-ATTTATTTA-TGC-A------CA-G-T---AGG......G-GCAGG-----CT-GA    134
CPZGAB        GATACA-TA---A-TGG.....-G--------GGTATA-----CATAGGCTTG-G-C-G----T--G--    131
CPZANT        TT--CTGTA-ACCTATACTCTATA-GCT-TAT----T---T-AG.........--GCAGC---------A    113

                                                                  env cds start ->
VPU_CONSENSUS ??ta???ttt?aaagAaTaAgaGAAagagca??????gaagacAGT???GgcaaTGaaAgt???gA?Gg?    159
A_U455        T%--...--AA-C-----------------.....---------...--------G---...--T--G    179
B_SF2         AT--...A--G-T-------------A----.....---------...------------...--A--G    176
D_ELI         TT--...C--G-T-------C----------.....---------...--------G--C...--G--G    177
F_BZ163A      GT--...-A-G--G-----------------.....---------...--------G---...--G--G    177
O_ANT70C      GAGG...--AAG--------------TTAGG.....--T--T---...-A-T-------C...A-T--A    189
O_MVP5180     AAG-...--AAGG-------AG----TCAGG.....--T-------...-A-T--------AAT--A-AA    192
CPZGAB        AC-T...AACCT--T-T-G-----------.....-----T---...--A---------...A-T--A    186
CPZANT        TAAGAGAAA-C--------C-T----GTGTT-AGTAGAAG-CTT---ATA-AT-G--C--TA...--A-AA    180

VPU_CONSENSUS GAtgaagAg???gaa????t??caga?cttgtgga?atg???gg??at?aTaatcct?gg??Tg?tgat?    207
A_U455        ---AC----...---...T-AT-CTTG--------G---...--GA-CT--G---T-G--GT--A-A--.    239
B_SF2         --CC-G---...---...T-AT---CA--------G---...--GC-CCT-GC----T--GA--T----.    236
D_ELI         ---AG----...A--...T-GT--A-A--G-----A---...--GC--C--GC----T--GA-AT----.    237
F_BZ163A      ----C----...---...C-GG---CA----G---A---...--GCC-TT--T----G--GA-AT--T-.    237
O_ANT70C      --A--G--ACAA---...G-TATG--C-----ACTT-GCCAT--CTT-G----C--CAT-TT--AGCC-    255
O_MVP5180     --AC--C--...---...G-CATG--G---A-AC-T-GCCAT--CTT-GC------CAT-TT--AGTTA    255
CPZGAB        --A-----A...--GAGAC-GGA-C-G---A-CC-T-AT...TACA--C----CAA-CATTT--C-A-CC    250
CPZANT        -----G--A...-CC...GATA--T-TTA-T---GGTCT...--CTT-GC------AGTTTA-AGA--AG    241

VPU_CONSENSUS ????????gA?ctg                                                         212
A_U455        ........A-TT--                                                          245
B_SF2         .........--T---                                                         242
D_ELI         .........--C---                                                         243
F_BZ163A      ........A-T---                                                          243
CPZGAB        CCATGTTT--C--C                                                          264
CPZANT        GG.......--CGAA                                                         249
```

HIV1 VPU CONSENSUS

```
CONSENSUS-A   ATgac?CCTTTg?????????gA?ATctGtGCAATAgtAGGaCTGaTAGTaGCaCTAATCtTAGCAATAG     59
CONSENSUS-B   ---caat----a.........c-a--A-ta----------c-t-AG--------gc---Aa---------     61
CONSENSUS-D   ---CAA-----A.........-tG--A-Ta----------C-t-Ag-------g-----Aa---------     61
CONSENSUS-F   ---T--?A?--?.........TTAGCAATAAG??--AC--C-T-A---------?----AA---------     54
CONSENSUS-0   ---CAT-A???-.........?-CC-GCTA--??--A--??T?-?-GT-CTTTG????-TA--AATG--?     46
CONSENSUS-U   ---CAA-----A.........ACA--AAC---------------AG--------T-C--AGC---C----     61
CONSENSUS-CPZ -----T????-A.........?T-G??-??????-TC-????T-????????T-?-????-GGA??--?T     27

CONSENSUS-A   T?GTGTGGACTATAGTAGgTATA...GAatatAAgAaA?tg?taaagcaaAga.................     106
CONSENSUS-B   -T--------c------ttc---..?---1----G----a-At---Ga------.................     111
CONSENSUS-D   -T---------------ttc---...--g-g--ga-g-t-aAa--g-------.................     111
CONSENSUS-F   -T-------------?-TA----...-------G----C--T---G-------.................     103
CONSENSUS-0   -T?-A---??G......TT--??......CT--GA-??TATT--G??----???AACAAGA?AGAA?GGA      90
CONSENSUS-U   -T-------GC------TA----...-------G----A-AAGG--A--G-AG.................     111
CONSENSUS-CPZ GTA-A?????????A-??T??A???G???----?A-??TATAA????--????.............?????     50

                                                                                -> env cds
CONSENSUS-A   ....AAAAtAgAcAggtta.??aTtaAgAGAATAAGaGAAAGAGCA...GAAGACAGTGgcAATGAgAGt     166
CONSENSUS-B   ..??---------------..----g-t-----------------...---------------------     171
CONSENSUS-D   ....---------t-----...---g-T-----------------?..-----t---------------     171
CONSENSUS-F   ....------A-T------...TA-G-A?----------?------...---------------------     161
CONSENSUS-0   G???G-----CTTGA?AG?...T-A-G?-------??----T?AGG...--T--?----A-T----A--?     143
CONSENSUS-U   ....----------------...C--G-T-----------------...-----T---------------     171
CONSENSUS-CPZ ?A?AG-?-?T?-G--A??T...?A???A-T??-???--????-???...??-?????-?-?-??--?---      79

CONSENSUS-A   GA?GGgGA?ACAGA?GAA...TTg?C?aaaCTt............GTgGAgATG...GGGaACTaTgATC     213
CONSENSUS-B   --a-----tcag--a---...--at-AGc----???????????---------???---C--c---c--     223
CONSENSUS-D   --g--?--t-g---G---...--at-agc----.....------------...---C--c---C--     222
CONSENSUS-F   --G-----TG----G---...?--G-A?C----.....-G---A---...---CCT-T-AT--     211
CONSENSUS-0   A-T-?A--AGA??-AC-?GAAG-?ATGG-?---.....?-AC?T-GCCAT--CTTTG?-A-?-     191
CONSENSUS-U   --T-----T-----A---...--AT-C-CT---.....A--------...---T--G-AT--A     223
CONSENSUS-CPZ ??-??A--AGA------G??A???GA-?????-????????????AT????????T???TTTGC-A-?-     107

CONSENSUS-A   tTgGGgtTGaTaAtAATtTGTAg                                              236
CONSENSUS-B   c-t---a--T-g--G--c-----                                              246
CONSENSUS-D   C-t---A-?T-G--G--a-----                                              244
CONSENSUS-F   C-----A-AT-G?----C----A                                              233
CONSENSUS-0   CCAT-...TT-G-G??????                                                 202
CONSENSUS-U   --TT--A-A--G----------A                                              246
CONSENSUS-CPZ C??T?...T?-??????????-CGAA                                           116
```

HMMER Sequences in the Env Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_HXB2R	HIVHXB2R	K03455	Starcich,B.	Science 227, 538 (1985)
C_UG268A2	HIVUG268A2	L22948	Louwagie,J.J.	JVI 69, 263 (1995)
D_ELI	HIVELI	K03454	Alizon,M.	Cell 46, 63 (1986)
E_TN2432	HIVTN2432	L03703	McCutchan,F.E.	ARHR 8, 1887 (1992)
F_BZ163A	HIV1BZ163A	L22085	Louwagie,J.J.	ARHR 10, 561 (1994)
G_LBV217	HIVLBV217	L11778	Louwagie,J.J.	AIDS 7, 769 (1993)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68, 1586 (1994)
O_MVP5180	HIVMVP5180	L20571	Gurtler,L.G.	JVI 68, 1581 (1994)
O_VAU	HIV1VAU	X80020	Charneau,P.	Virology 205, 247 (1994)
CPZANT	SIVCPZANT	U42720	Vanden Haesevelde,M.	Virology 221, 346 (1996)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
A_ROD	HIV2ROD	M15390	Clavel,F.	Nature 324, 691 (1986)
B_EHOA	HIV2EHOA	U27200	Rey-Cuille,M.A.	Virology 202, 471 (1994)
SD_MM251	SIVMM251	M19499	Franchini,G.	Nature 328, 539 (1994)
STM_STM	SIVSTM	M83293	Novembre,F.J.	Virology 186, 783 (1992)
VER_AGM3	SIVAGM3	M30931	Baier,M.	Virology 176, 216 (1990)
GRI_AGM677	SIVAGM677	M66437	Fomsgaard,A.	Virology 182, 397 (1991)
SYK_SYK	SIVSYK	L06042	Hirsch,V.M.	JVI 67, 1517 (1993)

The following alignment was generated using the HMMER program as described in the Introduction to this Part and in Part III. For simplicity, only representative PIV types and subtypes are shown. An ordinary consensus sequence (majority for lowercase, 50 percent or higher for uppercase) was deduced using MASE; this consensus is not a "most likely sequence" derived from an HMM model(see Part II). Annotation is based upon HIV1s, therefore the user should be cautious about its applicability to other PIV sequences.

```
ENV_CONSENSUS atGa?a???gtg????????????a??g???????g???aa????c??????t??????????????????    15
A_U455        ----G-...---.............-TG-GGATACA-AGG--TTAT-CTTGCT-GTGGAGATGG.......    48
B_HXB2R       ----G-...---.............-AG-AG..........--ATAT-AGCACT-GTGGAGATGG.......    39
C_UG268A2     ----G-...---.............-TG-GGATACA-AGG--TTGT-AACAATGGTGGATATGG.......    48
D_ELI         ----G-...-C-.............-GG-GGATAGA-AGA--TTGT-AAAACTGGTGGAAATGG.......    48
E_TN2432      ----G-...---.............-AG-AGACACA-ATG--TTGG-CAAACT-GTGGAAATGG.......    48
F_BZ163A      ----G-...---.............-GG-GGATGCA-AGG--CTGG-AGCACT-GGGGAAATGG.......    48
G_LBV217      ----G-...-C-.............-AG-GGATACT-AGG--TTGG-AACACT-GTGGAAATGG.......    48
O_ANT70C      ----T-...---ACTATGAAAGCA-TG-AG...AA-AGG--C...AAGAAGT-ATGG.............    48
O_MVP5180     ----C-...---ACT......ATG-AA-TAATGAA-AAG--CAACAGGAAGTCATGG.............    48
O_VAU         ----C-...-C-ATTATGAAAGCA-TG-GG...AA-AGG--C...AGGAAGT-AGGGATCTGG.......    54
CPZGAB        ----A-...--A.............-TG-AGAAGAA-AAG-GA.....................GACTGGA    34
CPZANT        ----GG...AA-.................................-CGATACATATTATTTGG.......    27
A_ROD         ----TG...AAT.................................-AGCTGC-TATT.............    21
B_EHOA        ---GC-CAT--T.............................--TAATTACCTAC-TGTT.............    30
SD_MM251      ---GG-...TGT.....................CTTGGG--T...-AGCTGC-TATC.............    30
STM_STM       ---GCC...TGC.....................CCTGGA--T...-AACTGC-TATC.............    30
VER_AGM3      ----AG...C--.................................ACATTAC-GATA.............    21
GRI_AGM677    ---GGG...AGA.................................TTGC-TATA.............    18
TAN_AGMB14    %%-CA-...-CA............TTTTGTATGAC-CCC--T.....................          25
SAB_SAB1C     ----AG...C-T.................................................          9
SYK_SYK       ---GC-...-CT.....................................................TTTAGAA    16

ENV_CONSENSUS ??????????????????????????????????????????gg?at??????????????t??t??t??g    23
A_U455        .....................................--A-CT.............A-GA-CT-GG-    65
B_HXB2R       .....................................--GTGGAGATGGGGCACCA-GC-CC-TG-    68
C_UG268A2     .....................................--A--C.............T-AGGCT-TT-    65
D_ELI         .....................................--C--C.............A-GC-CC-TG-    65
E_TN2432      .....................................--G-CT.............T-GA-CC-TG-    65
F_BZ163A      .....................................--CC-T.............T-AT-CC-GG-    65
G_LBV217      .....................................--A--T.............T-GA-CC-TG-    65
O_ANT70C      ...................................ACCT-G.............TACT-AGCCAT    65
O_MVP5180     ...................................A-CT-A.............TACA-AGCCAT    65
O_VAU         ...................................T-CT-G.............A-TT-GGCTTT    71
CPZGAB        ACAGC.....................TTATCCATAATT-CA.............A-CA-AACAAT    65
CPZANT        .....................................--TC-G.............GCTT-GC-AAT    44
A_ROD         .....................................-CC--T.............T-AT-AGCTA-    38
B_EHOA        ...........................................................ACAC-CCT    38
SD_MM251      .....................................-CC--C.............T-GC-TT-AA-    47
STM_STM       .....................................-CT--C.............T-GC-AT-AA-    47
VER_AGM3      .....................................--G--A.............C-AT-AA-AG-    38
GRI_AGM677    ....................................-AA.............A-AC-AA-AAT    32
TAN_AGMB14    ...........................................................          25
SAB_SAB1C     .....................................CTT-CA.............G-AC-GC-TT-    26
SYK_SYK       CTTACATAGTTTGTCTTTTTAGCTTAATAAGTTTAGGATTT...............................    57

                   signal peptide/gp120
                      <- vpU end
ENV_CONSENSUS g?t?tt??taat????tgta?????????????????a??tgt?tgtcacagt?tattatggggtacctgt?    65
A_U455        -T-GA-AA----T...----AT...GCA...CAAC-AT---GG-----G--C--C--------------G   126
B_HXB2R       -A-G--GA-G--C...----GT...GCTACAGAAA-AT---GG--------C----------------G   132
C_UG268A2     -A-A--AA-G--T...----AT...GTGATGGGGA-CT---GG--------C----------------G   129
D_ELI         -A-A--GA-G-CC...----GT...GCTGCAGACA-TC---GG--------T-----------G-----A   129
E_TN2432      -T-GG-GA----T...----GT...GCCTCAGACA-CT---GG--T-----T-----------T-----G   129
F_BZ163A      -A-A--GT----C...----AT...GCT...GAAA-CT---GG--------C----------------G   126
G_LBV217      -T-GG-AA----T...----AT...GCCTCAGGTA-CT---GG--------C----------------CG  129
O_ANT70C      -GCT--GA---CCCCA---TTG...AGCCTTAGAC-GC-A-A--CA-----C---GC------G-----A   132
O_MVP5180     -GCT--GC----CCCA---TTGAGTTATAGTAAAC-AT---A--C------T----C------------A   135
O_VAU         -A-AA-CCC-TGT...-TG-GC...TGT...AACC-AC-A-A--C------C----C------------A   132
CPZGAB        CA-T--GC---CCCCA---TTG...ACCTCTGAG...T-A-GG--A-----A--------A--------T   129
CPZANT        CCAG--TA--GAG...AAGGGG...ACGAATGAA...GAC-A---A-----A-TC-----A--C-----C   105
A_ROD         TGCT-GCT--G-A...-A-.....TGC...ACC...CAA-A---A--T--T-TC-----C-----CACG    93
B_EHOA        -C-TA-AAGT--C...-A-GGG...TATATGGGCA-GAAC-T------T--C-TC-----TA----C-CA   102
SD_MM251      TG-C-ATGGG--C...-A-.....TGT...ACT...CAA-A---------C-T------T-----A-CT   102
STM_STM       TGCT-GCT-G-CC...-A-.....TGC...ACT...CA--A------G--A-T------T-----A-CA   102
VER_AGM3      -A-AGGAG--G-G...CT--AT...ACA...AGGC-ACAA-GG--------A-T------A-----A--A    99
GRI_AGM677    AGCAA-AGGG--A...A---TA...GGAATAGGTA-CC---A---G-----G-T------AA-C--A--A    96
TAN_AGMB14    ...........................................................          25
SAB_SAB1C     -C-AAGTGGGTGT...--G-GCTTAGTGTGGTTAGTCCA--A---T-----G-T------TA-C--A--A    93
SYK_SYK       .....................ATGGAAAAAC-ACAA-A---A-----A-T-------A----ACAT   102
```

```
ENV_CONSENSUS tggaaagatgcaac?ac??c?ctattttgtgc?tcagatgctaa???a??????ac?ga????cataat?    115
A_U455        ------------GTT--CA-C-----------A-----------AGC-TATGATG-A--AGTG------G    196
B_HXB2R       -----G--A-----C--CA-T-----------A-----------AGC-TATGAT--A--GGTA------G    202
C_UG268A2     --------A-----A--TA-T-----------A-----------AGC-TATGAG--A--AGTG------G    199
D_ELI         -----G--A-----C--CA-T-----------A-----------ATC-TATGAA--A--GGCA------A    199
E_TN2432      ----G-------GAT--CA-C-----------A--------C--AGC-CATGAG--G--AGTG--C---G    199
F_BZ163A      --------------C--TA-T-----C-----A-----------ATC-TATGAG-GA--GGTA------G    196
G_LBV217      ---G----C---GAT--CA-G-----------A--T--------AGC-TATAGT--T--AAGC------G    199
O_ANT70C      ---G----------AC-AGTA-----C-----T-----------CCT-ACAAGC--T--AAAG------A    202
O_MVP5180     ---G----G---G-AC-AGTA-----C-----T-----------CCT-ACAAGC--T--ACAG------A    205
O_VAU         ---G---------AAC-AA-AT-G--C-----T-----------CTTGACAAGC--T--ACAG------A    202
CPZGAB        ---C-T-----TGACC-GGTA--C---------C-----C-----GGC-CATAGT--A--GGCT------A    199
CPZANT        ----G-A----G--AC-TA-T-----------CA--A----CTCCATGACAAGT--A--GGTG--C---G    175
A_ROD         ------A-------C-TTC-C--C---------AA-CAGA....--TAGG..............G--...A    142
B_EHOA        ------A-----T-A-TTC-C--C---------TA-CAGA....--CAG-..............G--...A    151
SD_MM251      ----GGA----G--A-TTC-C--C--C-----AA-CA-G....--TAGG..............G--...A    151
STM_STM       ----GGA----G--G-TTC-C--C--C-----AA-CA-G....--TAGG..............G--...A    151
VER_AGM3      ------A-CAGCT-AGTACAGGCT--C--CATGA--CCCA-C-CCAG-.....................C    148
GRI_AGM677    ------A--T----AGTTCAGGC------CATGA-GCCC....--TACC.....................---A    145
TAN_AGMB14    ..............................................................ACC---T     32
SAB_SAB1C     ------A--AGTT-AGTCCAAGCC------AAGA-CCCC....--TACG.....................--CC    142
SYK_SYK       ---G--------TATG-AC-T--C------A-CA-CTCCCA---AGG-.....................G    151

ENV_CONSENSUS t?TGGGc?aCacatgccTgtgt?cc???????acaga?ccc??????aatcca?aa???Gaa?ta???cT    160
A_U455        -C-----T-------------A--C.....-----C---.....--C---C--...---A--GATT-    251
B_HXB2R       -T-----C-------------A--C.....-----C---.....--C---C--...---G--GTAT-    257
C_UG268A2     -C-----T-------------C--C.....-----C---.....--C---C--...---A--GTT--    254
D_ELI         -C-----C-------------A--C.....--G--C---.....--C---C--...---A--GCA--    254
E_TN2432      -C-----C-------------A--C.....-----C---.....--C---C--...---A--TAC--    254
F_BZ163A      -C-----T-------------A--C.....-----T---.....------C--...---G--GTT--    251
G_LBV217      -C-----T-------------T--C.....-----C---.....------C--...---A--TCT--    254
O_ANT70C      -T-----AT----A--------T--T.....-----C---.....-C----T-T...---TATCCAT-    257
O_MVP5180     -T-----AT----A-----C--T--T.....-----T---.....------C-T...---T-TCCA--    260
O_VAU         -T-----A-----A--------T--C.....-----C---.....-G----A-T...---TATGAG--    257
CPZGAB        -T-----C-----G--A-----A--T.....-----T---.....-G---TC-G...---G--TTT--    254
CPZANT        -A-----A--TACCAGT-----G--A.....-T---T--A.....G----TATT...-T-G-TAGG--    230
A_ROD         CT----GA--CATACAG--C...TTG.....C-T--CAAT.....G--GATT-TCAG---A--ACTT-    197
B_EHOA        CC----GA--TGTACAA--C...-TC.....C----TAAT.....G--GACT-TACC---A-CCAAT-    206
SD_MM251      CT----GA---AC-CAG--C...-TA.....C----TAAT.....GG-GATT-TTCA---T-GGCC--    206
STM_STM       CT----GA---AC-CAA---...-TA.....C----CAAT.....GG-GATT-CTCA---T-GGCAA-    206
VER_AGM3      -A-----A--TAC-AA--CG...ATA.....C----TGAT.....C--GACT-CACA--GG--CCAT-    203
GRI_AGM677    -G-----A--CACCAA---CA-A--A.....GAT--T-AT.....G--AATAC-...--GG-GCCT--    200
TAN_AGMB14    -A-----C---ACAAAT---A-A--A.....GAT--T-AT.....---TACAC-...---G-GCAAT-     87
SAB_SAB1C     -A-----CT-TAC-AAT--CA-C--C.....GAT--TGAACCAGAAGGAA--AT-GCA---G--CCCA-    206
SYK_SYK       GA-----T--TA-GAAT-----G--GTCAGCA.................GATC--ATA---G--AGGG-    203

ENV_CONSENSUS ????AAtgt?aca???Gaaaa?TTt?a?at?Tgg??????aaa??????Aat?a????atg?????????    192
A_U455        GGTC-----G---...---G-A---A-C--G---.....---.....---A-C...---.....    294
B_HXB2R       GGTA-----G---...-----T---A-C--G---.....---.....---G-C...---.....    300
C_UG268A2     GGAA-----A---...----GT---A-C--G---.....---.....---G-C...---.....    297
D_ELI         GGAA-----G---...-----C---A-C--G---.....---.....---A-C...---.....    297
E_TN2432      GGAA-----A---...-----T---A-C--G---.....--C.....---A-C...---.....    297
F_BZ163A      GGAA-----A---...-----T---G-T--G---.....---.....---AGC...---.....    294
G_LBV217      GAAC-----A---...-----T---A-C--G---.....---.....---A-C...---.....    297
O_ANT70C      GCAC-----G---...--TG-C---A-T--A---.....---.....---T-C...---.....    300
O_MVP5180     AGGC-----G---...--T--C---G-T--A---.....---.....---T-C...---.....    303
O_VAU         AAAA-----G---...-GT--A--CA-T--A---.....---.....---T-T...--A.....    300
CPZGAB        TCCA-----A-T-...---TCA---A-C--G---.....---.....---A-T...---.....    297
CPZANT        C...---ACCT--...-TCTGG---A-TGCT-AT.....---.....---T-T...---.....    270
A_ROD         G...-----A---...--GGCT---G-TGCA---.....--T.....---.....-CA.....    234
B_EHOA        A...---A-A---...--GGCT---G-TGCA---.....G-T.....---.....-CA.....    243
SD_MM251      T...-----T---...----GC---G-TGCT---.....G-G.....---.....-CA.....    243
STM_STM       A...-----C---...--GGCT---G-TGCC---.....G-C.....---.....-CA.....    243
VER_AGM3      A...--CA-C--T...---CCA---G-AGCA---GCTGAC-G-.....--CCCC...T-A.....    249
GRI_AGM677    A...--CA-T---...---GCT--CG-GGCT---.....G-T.....---CCG...C--.....    240
TAN_AGMB14    A...-----CT--...--G--A---G-AGCG---.....---GACAGG--CCCT...T--.....    133
SAB_SAB1C     ACCT---A-C---...-----A---G-TGCT---.....---.....CGGAACCCT    252
SYK_SYK       C...--CA-T---GGA---T-C---CCAGCA---.....--T.....-GCAGTCAC---.....    249
```

```
ENV_CONSENSUS .???gTa???gaacaaatgca?ga?gAtaTaat?agttTaTggga?caaag?ctaAagCCaTGTGtaaAat  251
A_U455        ...---...--T--G-----T--A--------C---C-------C-----C----------------G-  358
B_HXB2R       ...---...-----G-----T--G--------C-----------T-----C----------------  364
C_UG268A2     ...--G...--T--G-----TC-G---G----C---C-------T-----C----------------G-  361
D_ELI         ...--G...-----G-----T--G--------C-----------T-----C-----A----------  361
E_TN2432      ...---...--G--G-----G--G---G----C-----------T-----T----------------G-  361
F_BZ163A      ...---...----------TACA--------C-----------T-----C----------G--G-  358
G_LBV217      ...---...-----G-----T--G--------C--C-------TG----C----------------G-  361
O_ANT70C      ...---...-----------G--A--C--T--T-----------C--G--T--T--A--T-----TC--A  364
O_MVP5180     ...--G...--C--------T--A--C--C--T-----G-----A--G--TT-------T----AG---A  367
O_VAU         ...---...--C--------C--A--C--T--AGA---G-----C--G--TT----A--T-----TC--A  364
CPZGAB        ...--G...--C--------T--A--C--T--C---C-C-----T--G--T--C-----T--------G-  361
CPZANT        ...---...---AGT---ACA--A-----GN-ACAA----TCC-A-----C-AT---------------C  334
A_ROD         ...---ACA------GCAATA--A---G-CTGGCA-C---TC--GAC-TCAA----A--------C---C  301
B_EHOA        ...--GACA--T--GGCAACAA-G---G-GTGG---C-C-TT--GACCTCAA----A--------T----  310
SD_MM251      ...--CACA-----GGCAATA--G--CG--TGGCAAC-C-TT--GACCTCAA-------T----------  310
STM_STM       ...--CACA------GCAATA--G---G-CTGG-ACC-T-TT--AAC-TCCA-T--A--T----------  310
VER_AGM3      ...---...-C----GCAGGAAGTA-----CACCTGC---TT--G--G-CT--G---------------  313
GRI_AGM677    ...---...A-----GCAG-GAGTA------CATCTAC-C-TT--A----CGA-G-G---T-----T--GC  304
TAN_AGMB14    ...--G...-C----GCAG-AAGCA-----CATCTGC-T-TT--AAGT-CCT-G-----T-----C--G-  197
SAB_SAB1C     CTA---...-GC---GCAG-GAGTA-C---CACCT---G-TT--AAGT-CTT-------------T--GC  319
SYK_SYK       ...A--AGGC-G-----ATTA--G--C--GTCAGC-----TCTTG---GCCAAT-GA--C-----G--GC  316

ENV_CONSENSUS TaaC?cc??T?TGtgTtacaaTgaa?TGta?taa?????a??????????????????????????????  280
A_U455        ----C--TC-C--C-----GT-AG-T--CCA---CATCACC.....................................  399
B_HXB2R       ----C--AC-C-------GTT-A--G--C-C-G-TTTGA-G.....................................  405
C_UG268A2     -G--C--AC-C---------TT-A--T----CA--TGTTA-T.....................................  402
D_ELI         ----C--AC-C-----C--TT-A--C----G-G-TGAATTGAGG...................................  405
E_TN2432      ----T--TC-C--C-----TT-A-TT----CC--TGCTA-GTTGACC.................................  408
F_BZ163A      ----C--AC-C--------TT-A--T----C---TGCCATT.......................................  399
G_LBV217      ----C--TC-C---------CT-A--C----C---TGCAATC......................................  402
O_ANT70C      -G--TTTCC-G-----ACA----G-G----CA--CATAGCT.......................................  405
O_MVP5180     -G--TTTCT-A-----ACA------C---GTAG-TCTGC-A.......................................  408
O_VAU         -G--TTTCT-G-----ACA------T----CAG-TATCA-A.......................................  405
CPZGAB        -G--C--TT-A-----A--CT--C-G--C-G---GGCTA-CTTTAGC.....................CA  410
CPZANT        ----A--TA-G---A-A-A------T----CAGGATAC..........................................  372
A_ROD         ----A--TT-A-----AG-------A--C-GC-GCACAG-GAGCAGCACAGGG.........AACAACAC  362
B_EHOA        ----C--AC-G-----G--------G----A---GACGTGGAGCTCA.................................  357
SD_MM251      --T-C--AT-A--CA----T----GA--C-A---AAGTG-GACAGATAGATGGGGA...TTGACAAAATC  377
STM_STM       -G--T--AC-A---A--G------GG----A---GAATG-AACAGACAAATGGGGATTGACAGGGAAAAC  380
VER_AGM3      --T-A--TT-G--CA---A----TCC---GTAG-ATTGA-CTCCTCTGAGCCTACC......ACCACTCC  377
GRI_AGM677    -CT-C--CA-A---A---A----TCC---GTAG-GCTG..........................................  342
TAN_AGMB14    ----A--AA-G--CA-A-G------T--C-C--GGCTTCCCTCCCCTACTCCC.........TCCTCCTC  258
SAB_SAB1C     -CT-A--AA-G--CA-A-A------C--CTA-CGCTTAG-AGGAGGA.................................  366
SYK_SYK       --G-A--CA-G---A---GG---TTG----CATTGGAC..........................................  354

ENV_CONSENSUS ???????Aa????ac????a?????ac?ac?a???c?ac????a?????????????????????????  294
A_U455        ....ATT--CAAC--C...-ATAAC-AC--C-ATATC--AGATGGAGTG...............................  441
B_HXB2R       ........--TGAT--T...-AT...--C-AT-GTAGT-GCGGG-GAATG...ATA.....................  444
C_UG268A2     ....GTT--TATT--C...-AT...-ATG-T-ATG-T--CAAT-GCCCCTACGAA.....................  447
D_ELI         ....AAC--TGGC--T...-TG...GGG-AC-ATGTC--TACA.....................................  438
E_TN2432      .AATGCT--TTTG--C...-AT...GTC-AT-ACATA--CAATGTCTCTAACATA.....................  456
F_BZ163A      ........--TGCC--T...GCC.....................................................  411
G_LBV217      ....GCT--TGTA--C...-AC...-GC-AT-AGA-T-GTAAC-ACAGCAGCCTT.....................  447
O_ANT70C      ....GGA-CA...--A...-AT.....................................................  417
O_MVP5180     ....ACA--TAAA--A...GGCCTATTA-ATGAGA-A-TAAAT.....................................  444
O_VAU         ........--TAGT-TT...-AT...--C--A-ACAGTC-CTTA-ACTCA.....................  441
CPZGAB        GGCAAAA--CCTA--A...-ACCAG--AT-TTCTC-GC-TCTC.....................................  450
CPZANT        ........--TGGA--A...CCT...--A--ACCAAGT--AACA-CAAGTACAGTAACACCAAAG.....  423
A_ROD         AACCTCA--GAGC--A...-GC...--A--C-CAA-C--ACCC-CAGACCAGGAGCAAGAGATA.....  420
B_EHOA        .GCTAGC--AGAG--C...-CT...--GT-CTCCG-GT-CCTC-GATCTTCTACTCAGACCCTACTC...  417
SD_MM251      ATCAACA-CAATA--AACAGCAGCAC-A--ATCAG-AC-AGTATCAGAAAAAATAGACATGGTC.....  441
STM_STM       AGTAACA-CAGTG--A...CCA...--AG-AGCAG-AG-AGCA-CAAAGCCAGAG...TTAGTA.....  435
VER_AGM3      TAAAAGT-CC...--G...GCCTCA--A--C-ATATC--AGCCTCAACAACCACTTTGCCGTGTGTGTCCAG  441
GRI_AGM677    ........--TGGT--A...GCC...--G--A-AGG-C--CACT-CTGCAACTACAACAATGACT.....  393
TAN_AGMB14    TAGTACA-CAAGG--A...-CA...-GA-ATCCCTGCC-AGGC-CAACA.....................  301
SAB_SAB1C     .GCAGCA-CAACA--A...TCA...C-AT-A-CAT-A--AGCACGCCCCGAGGTAGTGAGTGTG.....  423
SYK_SYK       ........--CAGCC-T...GCT...--A-GT-CCC-T--TACA-GTCCCCCTACTACACCTCCA.....  405
```

```
ENV_CONSENSUS ????????????????????????????????????????a??g?a??a???????????gaagaaa  305
A_U455        ..........................................................AGG.............-------  451
B_HXB2R       ..........................................................-TG-AGAA-.............-G---G-  460
C_UG268A2     ..............................AATGGGAAACTC-TG-A-CAG.............-G-----  475
D_ELI         ..............................................-AGGAG.............A---G--  451
E_TN2432      ....................................ATAGGA-ATAT-AC-.............--T---G  478
F_BZ163A      ..................ATT...CAAGACGGGGACCTGAAGGAA-G-CC-.............-GG-C--  448
G_LBV217      ........................GGGAATAGCACTGTG-GTAGCAT-.............-------  478
O_ANT70C      ..............................................GA-.............A-CCTT-  427
O_MVP5180     ..................................................................--G-  448
O_VAU         ..........................................-ACAATAC-.............A----GG  457
CPZGAB        ..................................................................----  454
CPZANT        ACAACAACACCA.....................................AT-GT-.............--T-GC-  448
A_ROD         AGTGAGGATACTCCATGCGCACGCGCAGACAACTGCTCAGGATTG-G-GAG.............-------  478
B_EHOA        AACGAAGATAGCAAATGTATTCAAAATGACAGCTGTGCAGGG-TA-G-CT-.............-------  475
SD_MM251      AATGAGACTAGTTCTTGTATAGCTCAGAATAATTGCACAGGCTTG-A-CA-.............--GC---  499
STM_STM       AATGAAACAAGCTCCTGTGTAAGTAACAACAACTGCACAGGCTTG-A-GA-.............---TC-C  493
VER_AGM3      AACAAGACAAGTACTGTGTTAGAATCATGTAATGAAACAATC-TA-A-AAGGAATTAAATGAA--GCCTG  511
GRI_AGM677    ACCCCCTGTCAGAATTGCAGTACAGAGCAGATAGAAGGAGAA-TG-C-GAG.............---CC-G  451
TAN_AGMB14    ......AACGATGAGTGCAATAGCCACTTGGTGACAAATAGC-TG-A-TTT.............---A-C-  353
SAB_SAB1C     GGCTTTAATGACTCAGTGATAGAACAAGAG.............-TG-A-AA-.............---C-GG  469
SYK_SYK       AATGAAACGTGG.....................TGGGGAGAC-ACAGTAC-.............--GCC--  442

ENV_CONSENSUS tga?a??????aa?TGt?c?TTtaatatgac?acag?a?taagaGataagaa??agaaa?a??a?gca?t  355
A_U455        ---A-......--T--CT-T--C--------C----A-C-------------AC-----GTAT-TT--C-  515
B_HXB2R       -A-A-......--C--CT-T--C-----C-GC---AGCA------G----GTGC-----G-AT-T---T-  524
C_UG268A2     -A-A-......--T--CT-T--C---G-A--C----A-A-------------AC---C-GCGC-T---C-  539
D_ELI         ---A-......--C--CT-T--C---G-A--C----T-C---A---------GC--C--GTAT-T---C-  515
E_TN2432      -A-G-......--C---T-T-----------C----A-C-------------GC----GGTCC-T---C-  542
F_BZ163A      -ACA-......--C---T-T--C---G----C-TG-A-G--%--------C-GC-----GTAC-T---C-  511
G_LBV217      ---A-......--C--CT-T--C-----A--C----A-G------------GTA-----G-AT-C--TC-  542
O_ANT70C      ---AG......--G---GAG------G-A--C--T-TTA-C-A---C--A--GG-----A-AC-G--TC-  491
O_MVP5180     ---G-......--T---AGT------G-A--T----TCC-C-C---C--A--GG--C--A-AC-G--TC-  512
O_VAU         ---A-......C-G---GAC------G-A--T----TGC-C-A---C--AC-GG-----A-AC-G--TC-  521
CPZGAB        ---A-......--C---T-A------G----A----A-T-------------GA-AC-GGTCT-TT--C-  518
CPZANT        ---AGCTACAAG-A---AAC------CA--GC----G-T-T-A---------AC-A---ATGA-A--CA-  518
A_ROD         C--TC......--T--CCAG--C--------AGG-TT-GA------------AA-AC-GT-TA-T-A-AC  542
B_EHOA        ---T-......G-C---CAA--C--A-----AGG-TT-AA-------G--TCAA-AC--T-TA-G-ACAC  539
SD_MM251      ---T-......-GC---AAA--C-CC-----AGGGTT-AA------C----CAA--G-GT-CA-T-A-AC  563
STM_STM       -AGT-......GGT---AAA--C--------AGG-CT-AA---------A--GAGAG-GT-CA-T-A-AC  557
VER_AGM3      CTTCT......--T---A-A---GCA---G-AGGGTATG--------C----AA----GT-TTCA-TGG-  575
GRI_AGM677    CATCC......--C--CA-T---GCA--TG-AGG-TATCA-------GTA--AA----TT-TAGCATGAC  515
TAN_AGMB14    GTTCT......-TG--CA-C---GCA---G-AGGGTACA-G--G---C----AA---C-T-TA-CAGCAC  417
SAB_SAB1C     CC-TG......--C---T-C---GC----G-AGG-TATAG---G---GTA--GA-A--TT-CTCAA--G-  533
SYK_SYK       GATTC......--T--CT-T-----CT----AGGG-GGT-C-AG-----A--AC--C--T-TAGA--TT-  506

ENV_CONSENSUS ?Tt?tAt??a???GAt?T?gt?????t????ga????aa?aa?????????????aata??Ag???????  380
A_U455        T--T---AG-CTT---A-A--ACAAA-C...A-TAAA-CTG-T.............--C-AT--T......  564
B_HXB2R       T--T---AA-CTT---A-AA-ACCAA-A...--T...--TG-T.............C--CC--C......  570
C_UG268A2     T--T---AA-CTT---G-A--ATCAC-T...--A...GGG--T.............G--AC-CT......  585
D_ELI         T--T---AG-CTT---A-A--ACCAA-A...--C...--TG-TAGTAGT...ACC----GT-CCAAT...  573
E_TN2432      T--T---AAGCTT---A-A--ACAAA-T...-GA...G-T--G.................----GT--TGAG...  591
F_BZ163A      G--T%--AG-CTT---A-A--ACCAA-T...AGC...--TGGAAGA.............---GGC--TGGAGAC  565
G_LBV217      T--C---AG-CTT---G-G--ACCAA-T...A-TGGT--TG-A.................----GT--TACAACT  597
O_ANT70C      A--C---GT-TCA---T-GA-GGAAC-G...A-TGAG-CA-GCAGCACAAATAAG-CA-AC--CAAA...  555
O_MVP5180     A--C---GT-TCA---C-GAGTAAGG-TAAT--C...TCA--TGCA......GTA---GGA-CAACA...  570
O_VAU         A--C---GTGACA---T-G--TAAGA-T...A-CGCC-CATCA.............---GAA-CAATG...  573
CPZGAB        A--T---GTGGAG---G-G--AAACC-A...-GG...--TG-G.............--C-AC-CA......  564
CPZANT        A--T---AA-GGA---C-TA-GAAGTGTCAG--C...--C--T.............G-G-CT-ACTGC...  570
A_ROD         A-GG--CTC-AAA---G-G--TTGTGAGACAA-T...--T-GCACA.............---CAG-CCCAG...  597
B_EHOA        C-GG---AA-CAA---T-A--GTGTGAAAAG-GGACA-GG-GT.................---GAA--CAAA...  594
SD_MM251      T-GG--CTCTACA---T-G--TTGTGAACAA-GG...--T-GCACT......GAT---GAA--CAGA...  621
STM_STM       G-GG--CTCTAGC---C-AA-CTGTGAACAAA-TGTC-CTGGC.............G-AGAG--CAGA...  612
VER_AGM3      G-GGA--GATGCA--AA-CA-GTGTAAGAAG-GT...--C--T.............TC--AC--AGAA...  627
GRI_AGM677    C-GG---GATCAG--GT-A--CTGCAAT...A-T...--A-CAGGAAGT...GAA--GGGA--TAAG...  573
TAN_AGMB14    A-GG---GATGCA--AC-CA-GTGTGAACCG--GAGC--A--G.............--CTCA--GGGG...  472
SAB_SAB1C     A-GGG--GATCAA--GG-G--TTGTGAAGAG-GAAGGG-A--ATCA.........--CGCA-CACATACT  594
SYK_SYK       C--C---AAGGAT--CT-GA-GAAAGAGGAA-GC.....................----GT--CTAT...  552
```

```
ENV_CONSENSUS ??????Tataa??TaataaatTGtAAtacctcagtcAT?a?acaagC?TGtccaAAg???tctTttga?c  435
A_U455        ......----GAC--------------------AC---T-C---G--T-----C---GTG--C-----G-  628
B_HXB2R       ......-----GT-G-C--G------C-----------T-C---G--C---------GTA--C-----G-  634
C_UG268A2     .......--C-GAT----------------------C---A-CG-----C---------GTCA----G--C-  649
D_ELI         ......----GGT----------------------C---T-C---G--T---------GTA--C-----G-  637
E_TN2432      ......----GGT----------------T--------T-AG--G--T---------ATA--C-----T-  655
F_BZ163A      ...TAC--C-GGC----------C---------AC---T-C---G--T--------AGTA----GG-GT-  632
G_LBV217      ...GAT----GGC--------------G-----AC---T-A------T--------GTGA-------C-  664
O_ANT70C      ...ATG----CAT---CT--------CT--A--AC---C-CG-----C---------GTA--------A-  622
O_MVP5180     .......----TGT---CT--------CT--A--A-T--C-AG--G--C-----G---GTAAG------G-  634
O_VAU         .......----GAT----T--------CT--A--AC---C-GG--G--C---------GTA--------G-  637
CPZGAB        .......----GGA----T-----C-----TA---C---A-C------A--------AACG--C-----G-  628
CPZANT        .......---T-CT--TGGC-C--C--C---A--ACT--C-C----T-C---GA----TCTA-------A-  634
A_ROD         .......-G-T-CA-G-ACC----C--C--A---------C-C-G--T-A---GAC---CAC-A--GG--TG  661
B_EHOA        .......-G-T-TA---A--CC---------A---A-T--CCA-G--T-A---GAC--ACAT-A--GG--TA  658
SD_MM251      .......-GCT-CA-G-ATC-C-----C--T--T--T--CCA-G--T-T---GAC--ACAT-A--GG--TA  685
STM_STM       .......-GCT-CA-G-G-C----C------AGT--A--TCA-G--T-C---GAC--ACAT-A--GG--TG  676
VER_AGM3      .......-G-T-TA-G--TC---------GAT-----T--A-A-G----C---GAT--AACA-A--GG--TG  691
GRI_AGM677    ...GAT-G-T-CA-G---C---------GAT-----G--A-A-G----T---GAT--AACA-A--GG--TA  640
TAN_AGMB14    .......-G-T-CA-G--CC--------CGAT--------A-AGG----T--CGA---AACT-AC-GG--TA  536
SAB_SAB1C     GTGGGA-G-T-CA-G---CC-----------------A--C-A-G-G--T---GAT--AACT-A--GG--TA  664
SYK_SYK       .......---T-CC-GT--C----------AAGT--T--ATCCGC---A---GAG--ACAGA-C---C-A-  616

ENV_CONSENSUS c?aTtcc?aTaca?TAtTGtgCtCCagC?GgttatgC?aT?cTaAaaTGtaatgatac?aa?TttaaaGG  497
A_U455        -A-----C-----T--------C-----T----T---G--T-----G-----G---C-TG-G--C--T--  698
B_HXB2R       -A-----C-----T--------C--G--T----T---G--T----------A---AG-CG--C--T--  704
C_UG268A2     -A-----T-----T----------T--------G--T-----G------A---AA-CA-----T--  719
D_ELI         -A-----C-----T--------C-----T----T---G--T-----G----GA----AG--G--C--T--  707
E_TN2432      -A-----T-----T------A-------T--------G--TT-T--G----------AG--T--C--T--  725
F_BZ163A      -A-----C-----T----------T--------G--TT----G----------AA--G--C--T--  702
G_LBV217      -A-----C-----T-----C--------T--G-T---G--T----C-----GG----AG--G--C--T--  734
O_ANT70C      -A-----A-----C----------A--A-----T--CT-T--G-----CAGC--AG-A-----T--  692
O_MVP5180     -C-----C-----C----------A-A--A-----C--CT-T--G--------C--AG-C-----T--  704
O_VAU         -C-----C-----C----------G--A-G---C--CT-T--G--------A--AGGA-----T--  707
CPZGAB        -T-----A-----C--------A-----A--G-T---A--T--T-----C-------AAG-C---TC---  698
CPZANT        -A-----T-----T----------A--A-----T--AT-G-G----G-A---GAGG-T----C---  704
A_ROD         -T--AAGGT-TAGA--C-----A---C-G--------CC-AT---G----------C--T-A-TC---  731
B_EHOA        GTT-AAGAT-TAGG----------CC-A--A-T---TT-G----G---C--------T--A-A-TC---  728
SD_MM251      -T---AGAT-TAGG--------A--TC-A--------TT-G--T-G---------C--A--T-A-TC---  755
STM_STM       -CT-AAGGT-TAGG--------A---C-T--------TT-G--C-G---C--------A--T-A--C---  746
VER_AGM3      AGT-AAGAT--AGG--C-----C--G--A--G-T---TT-AT-------C--C---TATG-T-A-GC---  761
GRI_AGM677    -TT-AAGAG--AGA--C-----A-----A--G-----TT-G---------------AGG-T-A--G---  710
TAN_AGMB14    -CC--AGGT--AGA--------A-----T-C-GT---T--CT-G--------A-----T--T-AC-C---  606
SAB_SAB1C     -CT--AGGC-TAGG----------A--------TT-G----G----GCA-----AG-T-A-TC---  734
SYK_SYK       -CT----G--T--A--C---------C-G-----CT-AT-GT-G-----C--------A--T---G----  686

ENV_CONSENSUS aa?a???gg???aTGta??AAtgT?a??gtagt?ca?tgtacacatggaat?aagccaac?gtaac?act  550
A_U455        --A-...--GCC---C-GG-----C-GCAC---A--A----------C------GTG---T-A---  765
B_HXB2R       --C-...--ACC-----CA-----C-GCAC---A--A----------T-G----GTA---T-A---  771
C_UG268A2     --C-...--ACC-----AT-----C-GCAC---A--A----------T--A---GTGA--T-A---  786
D_ELI         --C-...--CCC---C-CA-----C-GCAC---A--A----------T-G----GTG--GT-A---  774
E_TN2432      G-C-...--GCC-----AA-----C-GCTC---A--A--C----------T------GTG---T-A---  792
F_BZ163A      G-C-...--GCC---C-AG-----C%GCAC---A--G----------T--A---GTG---T-A---  768
G_LBV217      --C-...--ACCG----AA-----C-GTAC---G--A----------T------GTG---T-A---  801
O_ANT70C      --C-...--CAC---C-GA--CA-A-CG-----TACT----------C--T-G------A----GT---  759
O_MVP5180     --C-...--CCT---CCAC---A-TTCA--G--TACT----------C--C--------A----GT---  771
O_VAU         --C-...--TCTC----AA--C--T-CA-----TACT----------C--C--A-----A----GT--C  774
CPZGAB        --A-...--TAA-----CT-----A-GTAC---G--T----------T--A---GTG--T--C---  765
CPZANT        -GT-...--GATG----AA-----CTCA-----A--T--C--T--------A-GC----TG--GG-A--A  771
A_ROD         CTTTGCACCCAAC---TCT--A--AGTA-CTTCTACA--C--CAGG...--G-T-GA---GCA---TT-C  798
B_EHOA        CTTCATGCCCAAC----GT--G--AGTA---TCT...CTGTACAGA...--G-T-GA---ACAG--CT--  792
SD_MM251      CTTTATGCCTAA----TCT--G--GGTG--CTCTTCA--C---AGG...--G-T-GAG--ACAG--TT--  822
STM_STM       CTTTGCACCTAAT--CTCT--G--AGTA--CTCCTCG-----TAGA...--G-T-GAG--ACA---TT-A  813
VER_AGM3      GTTTAAGACAAAC---TCT-----TTCA--G--G--T-----TA-CTTG--A--TA----A--G--T---  831
GRI_AGM677    CTTTGCTCCAAAG--C-AG-----TTCA-----G--T-----TAGATT---C--TA-T--TA----T--A  780
TAN_AGMB14    CTTT...--CAC-----GA-----GTCA-----AAGT--C---GGCTT---G--TA-T--A----GCT-A  673
SAB_SAB1C     -CAT...AAGGC-----GG-----A-CA---TCAGCT--C--TAGGTT---C--TA-T--T----GC--A  801
SYK_SYK       -GAC...-ATGTC----CA-----A-CT-C---ATCA--C-----G.....................  729
```

```
ENV_CONSENSUS caa?t??t??t????aatgg?agt?tagcagaagaaaaaataa?aat?ag?t???aaaatat??caga?a  600
A_U455        ---C-GC-GT-A...-----C---C---------AG-G------GG--T--A-CTG-----T-CA--A-C-  832
B_HXB2R       ---C-GC-GT-A...-----C---C-----------G-GG--GT---T--A-CTGTC---T-CA-G--C-  838
C_UG268A2     ---C-AT-GT-A...-----T---A-----------G------T---T--A-CTG-----C-GA--A-T-  853
D_ELI         ---C-GC-GT-G...-----C---C-----------G-GG-C-T---T--A-CCG-----C-CA--A-C-  841
E_TN2432      ---T-GC-GT-A...-----C---C-----------G-G-----T---C--A-CTG-----C-CA----C-  859
F_BZ163A      ---T-GT-GT-A...-----C--CC----------G-T----T---C--A-CTC-------CT----T-  835
G_LBV217      ---T-AC-GC-G...-----C---T-----------G------TG--T--A-CTG----CT-CA--A-C-  868
H_VI557                                       A-CC-TG---T---T--A-CTA-------CA----C-   37
O_ANT70C      --GC-AA-AT-A...-----G-CAC-CT-TA---G--------G---G-TGGCAA--G----TTTG--AG  826
O_MVP5180     ---C-AA-AC-G...-----G-CAC-CT-TAG------G----G---T-TGGGAA-------TA----AT  838
O_VAU         ---C-AA-AC-A...-----G-CAC-CT-TA---G---T----C---C-TGGGAA-G-----TT----C-  841
CPZGAB        --GT-AC-TA-T...-----A---T----T----G---C----CTG-A--AGTGG-G----AGT--A-A-  832
CPZANT        TGGT-AC-AT-A...-----A-C-.....................................................  792
A_ROD         AC-TGGT-TGGCTTT-----C-C-AG------GA-T-G--C-TAT--CTAT-GGC-TGGC-GA...--T-  865
B_EHOA        AC-TGGT-TGGCTTC-----T-CAAGG-----A-C-GG-C-TAT--CTAT-GGC-TGG--AA...--C-  859
SD_MM251      ACTTGGT-TGGCTTT-----A-C-AG-------A-T-G--CTTAT--TTAC-GGC-TGG--GG...--T-  889
STM_STM       ACTTGGT-TGGCTTT-----A-C-AG-------A-T-G--C-TAT--CTAT-GGC-TGGC-GG...--T-  880
VER_AGM3      GG-C-GT-GT-G...-----G--CTACT----GA-TCG--CCCAG--AT-GCAGA--C---GAGT-AGC-  898
GRI_AGM677    GGGA-AGGAT-A...-----T---AG-T-----A-T-G--C-GAG--AT-GCAGA--GGAGGAAAT--T-  847
TAN_AGMB14    GCTT-TGGTC-A...-----T---CA------GA-T-G--C-GA---AT-GCAAA--C--GGAGT-AGT-  740
SAB_SAB1C     GG-A-AGGAA-T...-----T---TAT-T--C-A-T-G--C-GA---CT-GCAGA-----GGCAATTCC-  868
SYK_SYK       .................................---TTC-AT-C-T-GGCTAGCAC-TGGT-C.......  759

ENV_CONSENSUS atg??aa?a??at?atagt?ca?ct?aataaatct??????????????????????????????????  626
A_U455        ---CA--A-CC--A-----A--G--TGTC--TC--.....................................  867
B_HXB2R       ---CT--A-CC--A-----A--G--G--C-C----.....................................  873
C_UG268A2     ---CC--A-TA--A-----A--G--T--C------.....................................  888
D_ELI         ---CT--A-AC--A----CA--T--T---G-----.....................................  876
E_TN2432      ---CC--A-CC--A-----G--C--T---------.....................................  894
F_BZ163A      ---CA--A-CC--A-----A--CT-T---G-----.....................................  870
G_LBV217      ---CC--A-AC--A-----G--AT-T---------.....................................  903
H_VI557       --ACC--A-AC--A-----A--G--T--A-GTC--.....................................   72
O_ANT70C      G--GA--A-AT--C-----GACC--A--CTCTA-C.....................................  861
O_MVP5180     CA-CA--G-AT--C-----AACC--A--C-CTC--.....................................  873
O_VAU         G--GGG-G-AC--CC--A-AACC--A----CTAA-.....................................  876
CPZGAB        --ACTG-TGTCTGG-----A--G--AGTAG--G-A.....................................  867
CPZANT        ..............TAC--AACA--C-CT--AGTAGTAATGAATGGTCGCAAAAATGAATCTGTG..      843
A_ROD         --...-GA-CT--C--CAGC...T-A--C----A-TAT..................................  897
B_EHOA        --...-GG-CT--C---AGC...T-A---TC--A-TAT..................................  891
SD_MM251      --...-GG-CT--A--TAGT...T-A-----G-A-TAT..................................  921
STM_STM       --...-GG-CT--T--TAGC...T-G--C--G-A-TAC..................................  912
VER_AGM3      --...G-CTCAG-GT----GTTAT-T------CA-TAC..................................  933
GRI_AGM677    --...G-T-CAG-T---A-AA-GT-G-----G-T-TAC..................................  882
TAN_AGMB14    ---ACTCTGTC--T---AAG...--T-----GCA-TAC..................................  775
SAB_SAB1C     ---ATTCTGTT--T--TAGG...--C----G--A-TTT..................................  903
SYK_SYK       ..................--A--G---GGGA-CTACAAAGCAAAAGATAAAGTTAGATTTATCAAACA      809

ENV_CONSENSUS ??????????????????????????????????????????????????Taa?aaT?a??TGta?aAGa  641
A_U455        ..................................................G---A---C-AT---TCC---  888
B_HXB2R       ..................................................G--GA---T-AT----C----  894
C_UG268A2     ..................................................G--GA---T-AT---GCG---  909
D_ELI         ..................................................G---A---T-CC---GC---G  897
E_TN2432      ..................................................G--GG---C-AT----CC---  915
F_BZ163A      ..................................................G--CAG--T-AT----C----  891
G_LBV217      ..................................................A--GAT--TGTG----CC---  924
H_VI557       ..................................................G--CC---T-AC----CC---   93
O_ANT70C      ..................................................C---AC--G-CC---GA----  882
O_MVP5180     ..................................................A---AC--G-CC--C-T----  894
O_VAU         ..................................................A---C---AGCA---GAG---  897
CPZGAB        ..................................................G-T-GTC-A-AT---CAT---  888
CPZANT        ................CTTGTAAGATTTGGAAAAGAATTCGAAAACT---C---T-CA----T----  894
A_ROD         ..............................................AATC-C-GTT-GCAT----AG--G  921
B_EHOA        ..............................................AATC-G-C---GCAC----A---G  915
SD_MM251      ..............................................AATC---C---G-AA----G----  945
STM_STM       ..............................................AATT---C---G-GC----G----  936
VER_AGM3      ..............................................AATC---C-G-T-CT--C-A----  957
GRI_AGM677    ..............................................AACT-G-C-G-G-GA--CCG----  906
TAN_AGMB14    ..............................................AATC-G-C---ATTG----G----  799
SAB_SAB1C     ..............................................AATT---CT--C-GA----G----  927
SYK_SYK       AAAGGATAAAAATGAGTCAGTAATTATTTTAGTTCCAGAAGCACTTAGAT-GCAG--T-TA---GA----  879
```

```
ENV_CONSENSUS ccaggaaat?????????aa?aca?ta??a???ataa??ata??????gg??cagga?tagc?t??tat?   680
A_U455        --TTAC---ACAAGAAAA--T-T-CGT...AGGTAT-GT---......--AT-----CA---A-TC---G   949
B_HXB2R       --CAAC--C.........--T---AG-AA-AGA--CCGT--CCAGAGA--AC----GAG---A...-T-G   952
C_UG268A2     --CAAC---.........--T---AG-GA-AGT----GG---......--AC-----CA-A-A-TC---G   964
D_ELI         --CTATC-A.........--T---AG-CA-AGA-C-CCT---......--ACT---GCA-T-ACTC---A   952
E_TN2432      --CTCC--C.........--T---AGRCC-AGT----CTG-R......--AC-----CA--TA-TC---A   970
F_BZ163A      --CAAC--C.........--T---AG-AA-GGT---CAT---......--AC-----CG---A-TT---G   946
G_LBV217      --CAAC--C.........--T---AG-AA-AGT---CAC---......--AC-T---CA---A-TA---G   979
H_VI557       --CAAC---.........--T--GAG-AA-AGT----GT---......--GC-----CA---A-TC---G   148
O_ANT70C      ---CA--TA.........G-C-T-CA-...GAG--G-GA---......--T...CC-A-G--C-GG--CA   931
O_MVP5180     GA-----T-.........GCAGAGG--CA-GAT---TAT-C-......--TC--ATGAG-TGGCGCAG-A   949
O_VAU         ---------.........C-G---A--CA-AAG----TGGC-......--T...CC-A-G--T-GG--CA   949
CPZGAB        --------C.........--T---AGGGG-GAGG-GCAG---......--GC-----A-GA-C-TT---A   943
CPZANT        ---------.........-GG---G--AG-AATC--CAA---......--AC-----A-GA-T-TC---.   948
A_ROD         -----G---.........--G-T-G-GAA-CAA----TGC-T......ATGT-----CAT-TG-TTC-CT   976
B_EHOA        --------C.........--A-TGG-TGTGCCA----GA-CC......-TGT----TA-TCTC...-TC.   966
SD_MM251      ---------.........-G---G-TTT-CCAG-C-CC--T......ATGT-T---T-G-TT...-TC.   996
STM_STM       -----G--C.........--A---G-CTT-CCAG-C-CC--T......ATGT----CT-G-TC...-TC.   987
VER_AGM3      --------C.........--A---G-CTT-CCAG---CA--C......ATGG----GC---TG...-T-.  1008
GRI_AGM677    --T--T---.........--A---G-GTTGCCAG---CA--C......ATGG----GT---TA...-T-.   957
TAN_AGMB14    --------C.........--A---G-CTT-CCAG---CC--C......ATGG-----A-G-TG...-TC.   850
SAB_SAB1C     ---------.........--G---G--TTGCCTG---CC--C......ATGG-T--CT-G-TC...-TC.   978
SYK_SYK       ---------.........G-GT-TA-CAA-AAT--TCAAT-GGCTGCA--C..............---T   925

ENV_CONSENSUS ??a???????????????a????a?a?????????????????atagg????aa??taagacaaGC?ta   703
A_U455        TA-CA.........GGTA-A...-T-.....................-----G...G-TA----------AC-   986
B_HXB2R       TT-CA.........ATAGGA...-A-.....................-----A...--TA-G--------AC-   989
C_UG268A2     CA-CA.........GGAG-C...-T-.....................-----A...G-TA----------A--  1001
D_ELI         CT-CAAGA......TCAAGATCA-T-.....................-----A.............-----AC-   986
E_TN2432      GA-CA.........GGAG-C...-T-.....................-----A...G-TA----RAG---A--  1007
F_BZ163A      CA-CA.........GGAG-C...-T-.....................-----A...G-CA-C---A-G--CC-   983
G_LBV217      CA-CA.........GGCGCC...-T-.....................-----C...G-TA----------AC-  1016
H_VI557       CA-CA.........GGTG-C...-TC.....................-----A...G-TA-----A----A--   185
O_ANT70C      GC-TGGGA...ATAGGGGGA...-C-.....................GC---AAAC-GCTC---GGC---T--   977
O_MVP5180     TG-CACTTAAAAGAAGTA-C...-AT.....................-C-TCACCA-GATC---GGT---T--   998
O_VAU         GC-TGGCC...CTTAGTA-T...-C-.....................-AG--G...G-TAC---GGC---T--   992
CPZGAB        AT-TA.........GAAA-T...GT-.....................G----A...G-TACC---TCT--C--   980
CPZANT        ..-AC.........GTAG-A...-T-GCA...................-C---A...G-CACT--GA----G-T   986
A_ROD         CCCAC.........TACC-G...CCG.....................--CAATAAA-GACCC--------A-G  1016
B_EHOA        ..CAT.........TCAC-G...CCT.....................--CAATAAA-GACCT-A------T-G  1004
SD_MM251      ..CAC.........TCAC-A...CC-.....................--CAATGAT-GGCC--AG--G--A-G  1034
STM_STM       ..CAT.........TCGC-A...CCC.....................--TAATGAG-GACC--A------T-G  1025
VER_AGM3      ..CAT.........TCTC-G...-GG.....................TACAATACA-GGC-G--------T-G  1046
GRI_AGM677    ..CAC.........TCTC-G...-A-.....................TATAATACC-GGT---A------G-G   995
TAN_AGMB14    ..CAC.........TCAC-A...-A-.....................TACAATACA--AT-G-----G--A-G   888
SAB_SAB1C     ..CAT.........AGCC-G...-A-.....................TATAACACC-GGT-G-A------T-G  1016
SYK_SYK       TC%%C.........TCCC-G...TA-TACAAG%GGAAATTGAAA-CC--GAGGG-CGC--A--GT--C-T   980

ENV_CONSENSUS ?TGtaa?ttta?tgga????aaTGGaaa?aagc?ttaaAa?aggtaa???????????g?aaaa??????   745
A_U455        T-----TG-C-G-A--AGGG-C-----TAG-A-TA--C--C-----GCT.........-A-C--......  1041
B_HXB2R       T-----CA---G-A--GCAA-------TA-CA-T------C--A--GCT.........AGC---......  1044
C_UG268A2     C-----CA---G-A--AATG-------TAT-A-T---C--TG-----GA.........-A---......  1056
D_ELI         T-----TA---G-A--GCAC------GTA--A-T---C--C-A---GCT.........AG----......  1041
E_TN2432      T---G-GA---A----ACAA-------TAG--TT------C------CT.........-A----......  1062
F_BZ163A      T-----TG---G----ACAC-------TA--A-A---G--C-A----GG.........-C---G......  1038
G_LBV217      T-----TG---G--A-ACAG-C----G-G-CATG---C-GA------AA.........-C-C--......  1071
H_VI557       T-----TA---C-A-GGAAG-T------AGGA-T---C-TG-----GTT.........CAGC--......   240
O_ANT70C      T--C--G-A--A--CCACTG-T---GG-A--ATA------C-AAC-GCTGAAAGGTATTT-G--......  1041
O_MVP5180     T----CA-A--A-AAGACTGT----G--A-T--CC--C--C-AAC-GCTATAAGGTATTT---T......  1062
O_VAU         T-----T-A--G--CCACTG-C-----CA----C------A-CA---CTGAAAGATATTT-G--......  1056
CPZGAB        C-----GA-A-A---GACCAC------TCG-A-GG-GG--G-A----AA.........AAGGCT......  1035
CPZANT        C----CAG-C-A-AAGACGCT----G--C----ACGT--CA-AAC-GAG.........CACGTT......  1041
A_ROD         G--CTGG--C-AA--C...A--------G-C--CA-GC-GG----G-AGGAAACCCTT-C----......  1077
B_EHOA        G--CTGG----AA---...A-C----C-G----CA--C-GG----G-AAGAGACCATTAA---T......  1065
SD_MM251      G---TGG---GGA---...A-------GG-T--AA-----G----G-AACAGACCATT-TC---......  1095
STM_STM       G--CTGG---GGA---...G------G-GG---AA-----G----T-AGGAAACCTTG-TC---......  1086
VER_AGM3      G---C-C--CCAG--C...A-C----G-GG---C-GG---G-A----AAAATGAAATA-T----......  1107
GRI_AGM677    G--CC-C--CCAA---...G-T------GGG--A-GG---G-A--C-GAGAAGAAGTAAAG---GTGAAA  1062
TAN_AGMB14    G--CC-T--CCAA--G...G-C------GG---T-GG--GG----G-GAGAAACGATA-TT---......   949
SAB_SAB1C     G--CTGG---GGA--T...A-T----G-GG---C-GG--GG-A--C-AAGAAACAATA-TG-G-......  1077
SYK_SYK       C---CGAG-A-CA---...A-T----C-G--TTT--T--GC-----CAT.........-A-C-G......  1032
```

```
ENV_CONSENSUS ??????????????????????????????????ta?????????a?a?a??a?t??a???????????a    754
A_U455        ..................................T--..........-AGA-AA-G-TT-AT.........-   1060
B_HXB2R       ..................................T--..........-G-G-AC-A-TTGGAAAT.......-   1066
C_UG268A2     ..................................T--..........-A-AGGC-C-TCCCT.........-   1075
D_ELI         ..................................T--..........GG-ACCCTTCTT............-   1057
E_TN2432      ..................................T--..........-A-G-GC-C-TT-AT.........-   1081
F_BZ163A      ..................................T--..........-AGTCTC-T-TTCCT.........-   1057
G_LBV217      ..................................C--..........CA-GGAATC-AC............-   1087
H_VI557       ..................................T--..........-G-G-AC-C-TT-AT.........-    259
O_ANT70C      ..........................................................CTAGTA.........-   1048
O_MVP5180     ..................................C-T.........GT-A-CC-AACAGAG.........-   1081
O_VAU         ..................................C-T.........GT-G-AT-TAATCAAACT......G   1078
CPZGAB        ..................................T--..........GC-ACCTCA-CA............-   1051
CPZANT        ..................................C-T.........GCGG-GC-T-GG-AAAAAGTAGAC-   1069
A_ROD         ................................CATCCC.........-GGT-TAGAGGA-CC.........-   1099
B_EHOA        ................................CATCCC.........-G-T-TTCAGGA-CAACA.......-   1090
SD_MM251      ................................CATCCC.........-GGT-TACTGGA-CT.........-   1117
STM_STM       ................................CATCCC.........-GGT-TACAGGA-CC.........-   1108
VER_AGM3      ..................................T--CCAAAAGAT-G-T-CC-AGGA-CC.........-   1135
GRI_AGM677    AATCTTACAGAAGTAAGCATAGAAAATATACATC-G.........-G-AGGATA-GG.............    1110
TAN_AGMB14    ..................................T-GCCAAAACAT-A-T-TC-AGGG-CT.........-    977
SAB_SAB1C     ..................................C--CCACCAAAA-AGT-TAGTGGC-CA.........-   1105
SYK_SYK       ..................................GCC.........-CGA-AACC-GG-AA.........-   1051

ENV_CONSENSUS at?a?aca????????????ataa??tt??????a?cc??a???ggaggaGAtccaGAagt?acaa??c?    792
A_U455        -CA-A---.............----TC--TGCTAGCT--TC-...-----G---AT----A-T----CA-A   1115
B_HXB2R       --A-A---.............----TC--TAAGCA-T--TC-...-----G--C-------A-TGT--CG-A   1121
C_UG268A2     --A-A---.............----AC--TACTCA-C--TC-...-----G--CTT----A-T----CA-A   1130
D_ELI         -CA-A---ATA.........----AG--TAAACC-T--TC-...-----G--C-------A-T----CA-A   1115
E_TN2432      --A-G---.............----TC--TCAACC-C--TC-...----------TG---A-T----TG-A   1136
F_BZ163A      --GCA---.............----AA--TAATTC-T--TC-...-----G--C-T----A-T----TG-A   1112
G_LBV217      -CA-A-GC.............----CC--TAACTC-T--TC-...-----G--C-T----A-T---GCA-A   1142
H_VI557       --C-A---.............----TT--CGAACC-T--TC-...-----G--CATG---A-T----TG-A    314
O_ANT70C      -CA-T---GGTAGTATTAAC--G-CA--CAATCACAG-AGC...--T-------T---G--A--CCATTT   1115
O_MVP5180     --GTT--C.............----TA--CAGCAG-A-TAGT...--T------G-------A-GCCATTT   1136
O_VAU         --GTT--C.............--G-AA--CGGTAATCA-AGT...--T-A----G-------A----ATTT   1133
CPZGAB        --CGA---GCAGCC...AAC----CAC-CAACAG-G--TCT...--------------G--C---CAT-A   1115
CPZANT        -CA-A--CAATGCGAAAACA---TGGACATTCCA-GAT.........--------T-----A-A-GTG-A   1130
A_ROD         --G-C---AGG......AAT--T-GC--TGCAGCGC-AGG-AAA--CTC---C---------AG--TACAT   1163
B_EHOA        --ATCT--CAG.............----GG--AGCAGAGCA-GCGAGAA-CTC------------A-G-TATAT   1151
SD_MM251      -CA-T--TGAT......AAA--C-AT--AACGGCTC-TGG-...-----------G-----T--CTTCAT   1178
STM_STM       --G-C--GGCA......AAA----GGA-AGTGGCTC-TGGG...--------------G--C--CTTCAT   1169
VER_AGM3      --G-T--TGAA.........GAG-TT-ATCTGCAGAGACT-...TTT-------------CAG---ATTT   1193
GRI_AGM677    ..........................................------------TCAG-G-ATTT   1133
TAN_AGMB14    -CG-T--TAAG......AAA---TGGC-AAGGAG-CAA.....T-G-----C-----G-CAG-G-GCAT   1035
SAB_SAB1C     --G-C---AAC.........-AG-TT--TCTACA-AGACA-...T-G------------TCTGA-TTCTT   1163
SYK_SYK       --GTA--C.............-AC-CCACATGGAGGT--CAGCCA--G-------T---G--ACG--CA-A   1109

ENV_CONSENSUS ?tg?tttaAtTGt?gaGGaGAaTTttTcTAtTGtaataca?ct?gg?tgtTtaat?????ta????????    843
A_U455        TA-C---------G-------------------C--C---T-AG-CC--------...AGC-TTTGG...   1179
B_HXB2R       CA-T---------G----G---------C------T--A-ACAAC--------...AG--CTTGGTTT   1188
C_UG268A2     TA-C---------A-------------------C------T-AA-CC--------...AG--GTGATAAT   1197
D_ELI         CA-T---------G----G-----C-----C---------T-AG-AC--------...AG--CATGG...   1179
E_TN2432      TCAT---------A----G--------------C------A-AC-AC--------...AA--CTTGCATA   1203
F_BZ163A      TA-T---------A----------------C--C------T-AG-AC-------CGAC.............   1170
G_LBV217      TA-T---------A-----------------------T-AG-AC---------...AG--GTATATTA   1209
H_VI557       TACT---------A-------------------C------T-AAAAT-------C...AG--CTTGGGTT    381
O_ANT70C      ACAC-----C---CAT--------C--T--------C---G--AA-A--------...TA--CCTTTTCA   1182
O_MVP5180     ACAT-----C---CAT--------C--T--------C---T--G--A-------C...TA--CTTTTATC   1203
O_VAU         C-TT-----C---CAT--------C--T--------C---AA-C--C--------...CA--CCTTTTCC   1200
CPZGAB        TATG---------G-----------------------TAGCCAAA-A----C-GAC.............   1173
CPZANT        T--G--------CCA---------C--T------G---T-A-ACCTTG---C---...GCC-CATACACG   1197
A_ROD         G--GAC---C--CA-------G---C----C--C--C-TGA--T--T-CC-C---...TGG-TAGAG...   1227
B_EHOA        G--GAC---C---A-G-----G---C----C-------TGA--TTTT-C--A--C...TGGGTAGAA...   1215
SD_MM251      G--GACA-----CA-------G--CC----C-----A-TGAA-T--T-TC-A---...TGGGTAGAGGAT   1245
STM_STM       G--GACA-----CC-T---------C-T-----C--A-TGAA-T--T-C--A---...TGG-TAGAAAAC   1236
VER_AGM3      A--G---------CAG--G-----C-----------A-TGGA-T--T-TC-A---...TACCTGAAT...   1257
GRI_AGM677    T--G--C------CA---T----------------G-TGGACT--T-TA-C---...TA-CTAAAC...   1197
TAN_AGMB14    A--G--------CCTG-------------------CAC-TGA-T--T-TC-A---...TA-TTGAAC...   1099
SAB_SAB1C     C-TT---------CA---T-----C----------A-TGGA-T--T-CC-C--C...TA-TTGAAT...   1227
SYK_SYK       T--G--CC-A---G-G-----G--C-----C------GT-T--AA-C-C---GC-...AAC-TTACG...   1173
```

```
ENV_CONSENSUS ?????????aatag?ac??????????????????????????????????????????aat???????   853
A_U455        .........---G-C-GCATGTCA.....................................---GAC....  1200
B_HXB2R       .........-----T--TTGGAGTACTGAA.....................GGGTCA---AAC....  1221
C_UG268A2     ......AAT-----T--C.....................................................  1209
D_ELI         .........----TT-GTGCATGG.....................................---AATATT.  1203
E_TN2432      ......GGA---GAA--CATG.....................................---.........  1221
F_BZ163A      ..............--A.....................................................  1173
G_LBV217      ......TCT--C-AT-ATGCACCA.....................................TC-.....  1230
H_VI557       .........--C--T--ATCA.....................................---GATATC.  402
O_ANT70C      ......TGT--CG-A--CACCTGTAGT.....................................GTT-G-AAT....  1212
O_MVP5180     .........--CT-T--AAAGTCCGGATGCCAGGAG.....ATCAAAGGGAGC---.......  1245
O_VAU         TGCAAGAAG----TG--CAATAACAAGATC.....AATTGTACT---ATTAGC.  1248
CPZGAB        .........--C-TT--A.....................................................  1182
CPZANT        ......GGA--CCTC-TC.....................................................  1209
A_ROD         .........----AG--A.....................................................  1236
B_EHOA        .........-----A--T.....................................................  1224
SD_MM251      ......AGGG--GTA--TACCCAGAGG.....CCA--G.....  1272
STM_STM       .........-GA--T--ATCAGAAATGAGAGAT.....TGGAACAAA--CAAA....  1275
VER_AGM3      .........---C-T--AGTAGATCCGGACCAT.....AATCCGTGT---GGT....  1296
GRI_AGM677    .........---C-A--AGAAGAT.....GCAG-AGGT....  1221
TAN_AGMB14    .........---GAGT-CAGTGAAGGAAGCTTCACTGACGTAGAAGGCAATAGATGTTCA--CATAACAA  1160
SAB_SAB1C     .........----AGT-AGTAGACCCAGACCATAAC.....AACTGTGCAAAA---AAT....  1272
SYK_SYK       .........---G-A-ACGCAAGT.....AAG---AAT....  1197

ENV_CONSENSUS ??????????????ga??????????????????????????????????a?t?a?a??at?a?   862
A_U455        ...........ATG-GGCCA.....-A-GGC-CT--A-C  1223
B_HXB2R       ...........ACT--AGGA.....-G-G-C-CA--C-C  1244
C_UG268A2     .....-TC--C-C  1217
D_ELI         ...........ACA--GTCAAATAAT.....AGCACA-ACACA-AC--C-C  1238
E_TN2432      ..............-GGTGT.....-A-GGC-CT--C-C  1241
F_BZ163A      ..............-GATCC.....-A-GGC-CT--C-C  1193
G_LBV217      ..............A-TAGC.....-G-A-G-AT--C-C  1250
H_VI557       ...........TCAA-TGGC.....-CAACA-TA--TGA  425
O_ANT70C      ...........GTTAGTCAA.....GG-A-C-ATGGC-C  1235
O_MVP5180     ..............--GACC.....-A-A-A-ATGGT-C  1265
O_VAU         ...........AATA-TAGC.....-A-GGC-CTCAGGC  1271
CPZGAB        .....-A-GGC-TT--A-T  1196
CPZANT        ..................ACA.....-ACGGAGCCC-C-T  1226
A_ROD         .....C-CCGC-ATTA  1247
B_EHOA        ..............-GG.....CTCA-G-GA-ATTA  1241
SD_MM251      ..............--ACGG.....CA-AGA-GG-ATTA  1292
STM_STM       ...........AAG--ACAA.....CAAA-G-GG-ATTA  1298
VER_AGM3      ...........ACGA-GGGAAAAGGTAAGGCA.....CCAGGACCCTGTGCAC-A-GA-CATA  1343
GRI_AGM677    ...........ACTA-TAGGACCTGTGACAAAGGGAAGCCAGGACCAGGACCATGTGT-C-G-GA-CTTA  1280
TAN_AGMB14    GTGGAGGACTGACA-GGTCCACCAGA.....AAGTGTCT-A-A-GG-CCTA  1206
SAB_SAB1C     ...........ACAA-GCCATGT.....TGGC-G-GA-CCTA  1298
SYK_SYK       ...........TAT-CC.....AGT-ACC-TCG  1214

ENV_CONSENSUS tct?ccaTgcaa?aTaagacaaaTa?TaAa?a??tgGcaga??gtaggaaaag?aat?TAtgc?cCtCCc   921
A_U455        ---C-A---T-GA----AG-----TA----T-TG-------GA------C---C---G-----C------  1293
B_HXB2R       C--C-------GA----A------TA----C-TG-------AA----------C---G-----C------  1314
C_UG268A2     A--C-------GA----A------TA----C-TG------GGG------CG--C---G-----C------  1287
D_ELI         A--C-A-----GA----A------TA----G-TG%%-TG-CAG-C----%---C---A-----C-----T  1305
E_TN2432      A--T--------G----AG-----TA----C-TG------GGA-C----C---C---G-----T------  1311
F_BZ163A      ---C-T---TCGA----A------GG----C-TG------GAA--G---CG--C---G-----CG-----  1263
G_LBV217      A--C-----T-GA----A------TG---GA-TG-------GG--G---C---C---G-----C------  1320
H_VI557       A--G-------GA----A------TA----C-TG-------GA------C---C---G-----C------  495
O_ANT70C      ---A--T-----AC-G-----GG-GG---GGTCA---ATA-GG-G-CAGTCG-G-C-C-----A------  1305
O_MVP5180     -A-A--T-----GT-------GC--G---GATCA---AT--AG-G--AGTCGAG---C-----A------  1335
O_VAU         AA-A--T----GGT-G------G--G---GGGAC---AT--GG-G----TCG-G-C-T-----A------  1341
CPZGAB        A--G-------GA--T-----G---G-T-GTTCA---AT--GG-------G--G---A-----A--C--A  1266
CPZANT        AGCA-AT----GA--T-AG--G---G-T--TCAT---GGC-TA--TTC-----GC--T--CTTAG-C--T  1296
A_ROD         -GCA--G---C-T----AG------A-T--C-CA-----T-AG-----G-G-AATG-A---TTG------  1317
B_EHOA        -GCGT-----C-C--C-----G---G-C--C-CG-----C-AGA-T----G-AATG-G---TTG-----A  1311
SD_MM251      CG-G--G--TC-T--T---------A-C--C-CT-----T-AA-----C---AATG-T---TTG-----A  1362
STM_STM       -G-A------C-C--C------G--A-T--T-CA-----T-AA-----G---AATG-T---TTG-----A  1368
VER_AGM3      -G-TG-T---C-T---C--TCTG-CA-T--TGAT---T-C-CAC--TC--GGAA--CC-----A--G--A  1413
GRI_AGM677    -G-TG-C---C-T---C-----G--G----TGAT---T-C-CT--CTCT---AAGG-A-----T--A--A  1350
TAN_AGMB14    CG-TGGC-TGC-T---C-GTC-G-TG-C--TGAC---T-T-CAC--TTGC--AA--GA-----G--G--A  1276
SAB_SAB1C     -G-T--C---C-C-----G---G--G----TGAT---T-C-CCT-GTCT---AAG-CA--C--C------  1368
SYK_SYK       ---GT----TGCC--TC-C------A-T--TGAT---AGATAT---A-----TTG--A--CCTG--C--T  1284
```

```
ENV_CONSENSUS At??aaggaaac?T?ac?TGtaa?TCaaa?at?AC?ggacTa?ta?t?a?aatagat?g???t???aat?  970
A_U455        --CC-----GTAA-A-GG---G-A-----C--T--A------C--T-A-C--G----G-GGG-...-C-.  1359
B_HXB2R       --CAGT---C-AA-T-GA---TCA-----T--T--A--G--GC--T-A-C--G----G-TGG-.....-A  1381
C_UG268A2     --TA-------AA-A--A----GA-----T--C--A------C--C-G-C-CGT---G-AGGGGAG-CA.  1356
D_ELI         --CG--A-----A-TCTA---TCA-----T--T--A--G---C--T-G-C--G----G-TGG-ATA---.  1374
E_TN2432      --CAGT-----AA-T-AT---GTA-----T--T--A---A--C--T-G-C--G----G-TGG-GCT---.  1380
F_BZ163A      --CTC--A----A-T--C-----C--G--T--T--A--T---C--T-G-C--G----G-TGG-CAG---A  1333
G_LBV217      --TGC-------A-T--A-----A-----C--T--A------A--T-A-C--G----G-TGGG...---A  1387
H_VI557       --CA-------AA-T--A---G-A-----T--T--A--G---A--T-G-C--GT---CATGT-...---T   562
O_ANT70C      --CA----T--TC-A--A----TG-----C--A--T---A-GA-CC-ACA---G---AACACA...TGG.  1371
O_MVP5180     --CCCC--C---T-A--A---C-T--C--C--A--T---A-GA-TC-ACAGT-----CAACCA...TGG.  1401
O_VAU         --CCC-------C-AGTA--C-GG-----C--A--T---A-GA-TC-ACA-T-G--CACGCCATGG---A  1411
CPZGAB        --CAG-------A-C--T--C--T--T--C--T--T-----TC-CT-A-CTTC---CACTCC-GTG-C-.  1335
CPZANT        -GGAG---G--TG-TT-C----CT--C-GC--A--T---A-TA-GT-GGA-.........GG-CAA-TAT  1357
A_ROD         -GGG----GG-GC-GT-C--C--C-----CAG-A--CA-CA--A-TGCT-AC--T--CT-GCAAAAC---.  1386
B_EHOA        -GAG-G--TG-AC-CT-C-----T--C-CTG-T--CA-C--CA--GCC-AC--T--CT-GATAGAT--G.  1380
SD_MM251      -GAG-G---G--C-C--G-----C--C-CAG-G--CA-T--CA--GCA-AC------T-GAC-GATGGA.  1431
STM_STM       -GAC-----G--C-T--T-----T--C-CTG-G--CA-TA--A--GCA-AC--T--CT-GAC-AAC---.  1437
VER_AGM3      -GAG----GC--T-GCAA--C-CA--C-CGG-A--G--TA-GTC-G-GGAGC--A--TACAA-AGT--G.  1482
GRI_AGM677    -GGG----TC-TT-GGAG-----C---TCAG-C--G-C----TACG-GGC-------TATAAC....-C.  1416
TAN_AGMB14    -GAG-----C--T-GGAG--C-GA--C-CGG-G--ATC-A-GC-GG-GTCGC-G---TATAACGAC--G.  1345
SAB_SAB1C     -GAG----CC--T-GGAG-----C--G-CAGCC--A-CG--GTATG-AGA-T-GA-CTATAACAGT--A.  1437
SYK_SYK       -CGGC---GC-TA-T-AA----CT--C--TG-C--T-C-G--T-GACGGAC-----ATACTACCCAGGA.  1353

ENV_CONSENSUS ???????????aacaa?ac?aa???????gt?acctt?agacCt??agg?ggaga?atga??gacaattg  1012
A_U455        ..........-----T--A--A...AAT-AG-----C------GG---A-----T----GG--------  1415
B_HXB2R       GC.........-----TGAGTCC.....-AG-T---C------GG---A-----T----GG--------  1436
C_UG268A2     ..........-GTG-G--A--TAGCACA-AG--A--C------GG---A-----T----GG--------  1415
D_ELI         ..........--T-GT--T--C.....-AG-----T------GG---A-----T----GG--------  1427
E_TN2432      ..........-CT-CG--T--C.....-AG-----C------GG---A---A-T--A-AG--------  1433
F_BZ163A      AT.........--TC-G--T.........AG-----C------GGG--T--T--T----AA---T----  1385
G_LBV217      GT.........-C---TGGA--T.....-AG-----C------GC---A-----C----GG--------  1442
H_VI557       CT.........GCAG-A-GT--T.....--T-----C------GG---A-----T----GG--------   617
O_ANT70C      ..........----GC-GC--CAACAAT--A--A--T-----AAT---G-----C----AA--T-TA--  1430
O_MVP5180     ..........--TTCC--AGGT...GAAAAT--AC-T-----AGT---G-----T----AA--T-TA--  1457
O_VAU         AAACACATCCT----GC--C.....----C-T-----AGG---G-----T----AA--T-TA--      1469
CPZGAB        ..........-----C-GTGGT...AACC-G--A--T------AC---A---A-C----AA--T-T---  1391
CPZANT        AT.........--TG-A--T.........--T-AAG-GTC----...-CT-C-AGAG-AGCA---C-A--  1406
A_ROD         ..........--TC-G--A--C......A-T-----T--TG-A...-AG-TG-CAGAA...CTAT-C..  1431
B_EHOA        ..........---CTT--T--T......A-T--TG-G--TG-A...-AA-TGTCAGAA...CTGT--..  1425
SD_MM251      ..........---C-A--T-GT......A-C---A-G--TG-A...-AG-TG-CAGAA...CTGT--..  1476
STM_STM       ..........--TG-G--T--C......A-C--TGCA--TG-A...-AG-TG-CAGAA...CTGT--..  1482
VER_AGM3      ..........----GG--T--T.....--A--A--A--T--C.....C-G--AGAAAC--TC--       1526
GRI_AGM677    ..........--GTCTGGCCCAATAAAT--G---C-A--T---.....C-GG-ACGCAG--TA--     1466
TAN_AGMB14    ..........----TG--A--T.....--C--A--A-C-G-A.....A-TT--GAAA---TA--      1389
SAB_SAB1C     ..........----GG--C--T.....--G--G--ATC----.....C-G--TGAAAG--T---      1481
SYK_SYK       ..........TCT-CGTTG--C.....--T-CT--AAC--CAAAT-TGGA-...--TGT---       1397

ENV_CONSENSUS gagaa?tgaatTgt??aa?TAtaAagTagTagaaaT?aaaCCa?TtggagTaGCaCCcACaaa?g?aaaa  1074
A_U455        --A--G------A-AT--G----------A----TG-----C-A--------------C-GG-C---G  1485
B_HXB2R       -----G------A-AT--A---------A----TG-----T-A--------------C--G-C---G  1506
C_UG268A2     -----G------A-AT--A--------G------G-T--G---C-A--------------T--G-C---  1485
D_ELI         -----G------A-AT--A-----G------C----TG-----C-A--------------C-GG-C---G  1497
E_TN2432      -----G------A-AT--A----------C----TG-----C-A---A----------C-GG-C---G  1503
F_BZ163A      -----G--T---A-AT--G--------------TG-----C-A--------------C-GG-C----   1455
G_LBV217      -----G------A-AT--G---------A--A----T------C-A--------------C--G-C--GG  1512
H_VI557       -----G------A-AT--A----------AG-G-CG-----C-A---T----------C-GG-C--GG   687
O_ANT70C      -----C--------TC--C--C---------AGGG-A-----TT--A-T--G----------CGTATTGC-  1500
O_MVP5180     -----C-A------AC--C--C---------C-G--A-----TT--A-T--------T-----AATGTC-  1527
O_VAU         -----C-C------TC--A----------AG-G-A-----TT--A-T---------A-----AATTGC-  1539
CPZGAB        -----G---GC-C-AT--G----------TCG---AG-G--TT-ATC---G-----A-----A-C--G-  1461
CPZANT        ----GCG--G----CC-GG--CC-G--G-----G--TGRT--CT-GTC------C--A----CA-GN---  1476
A_ROD         .---TTG--G---GGAG-T------T-G--------A-C----A----CT-C-----T-----A-A----  1500
B_EHOA        .-A-TTG---C--GGGG-C--C---T-------G--A-C----A----CT-T-----T----GTAT----  1494
SD_MM251      .C--TTG--G---GGAG-T------T-------G--C-CT--GA----CT-G--C------G-T-TG--G  1545
STM_STM       .C--TTG------GGAG-T--C---T----------A-CT---A----CT-G--C--A--G--T-TG--G  1551
VER_AGM3      -GC-GCA------GGC-GG--C---T----------T-C----A----CT-C----------G-A-T--G-  1596
GRI_AGM677    -GCGTAC---C--GGAG-C------T-------G--A-C----A----CT-T--T--T--T----G-T-T--G-  1536
TAN_AGMB14    -GCTTA-------GGC-GG-----GC-CA-T--G--C------A-A--GT-T--T--A---G-G-TG-GG  1459
SAB_SAB1C     -GCT-A-------GGAG-T--C---T----G-----C--G---A-----T-T--T--A--C--A-T--G-  1551
SYK_SYK       ----GCA--TC---TC--T------T-GA-TC----T------A----CT-T--G-----TG-TCA-CGT  1467
```

```
                                                                gp120 / gp41 in HIV1s
ENV_CONSENSUS aGa??ag??gt?g??????????????????AGagaaAaaaGaGcagt???????ggactaGga????t?t  1110
A_U455        ---AG--TG--G-AG.....------------------T.....---------GCTA-C-              1534
B_HXB2R       ---AG--TG--GCAG.....-----------------G.....---A-----GCTT-G-              1555
C_UG268A2     --GAG--TG--G-AG.....-----------------G.....---A-----GCTG-G-              1534
D_ELI         ---AG--TG--G-AA.....---------------A-A.....---T-----GCTA-G-              1546
E_TN2432      ---AG--TG--G-AG.....-----------------G.....---A-----GCTA-GA              1552
F_BZ163A      ---CA--TG--GAGG.....-----------------G.....---A-G---GCTT-G-              1504
G_LBV217      ---AG--TG--G-GG.....---------------A-T.....ACT---GCTG-C-                 1558
H_VI557       ---AG--TG--G-AA.....--------------T--G.....-----G-A-GCTG-T-               736
O_ANT70C      --GCC--TCA-AAGCACTAGA...ACTCAT-----------------A.....---T-G---ATGC-A-     1561
O_MVP5180     ---CC-ATAA-AAACATTCACACCCCTCAC--G--------------A.....---T-G---ATGC-A-     1591
O_VAU         --GCC-ACTA-A-GAACTAGA...TCTCAT-----G----------CA.....--TT-G-C-ATGC-A-     1600
CPZGAB        --GCATACA--G-CAAGACAG...AAAGAC---C-G----------CCTTC...--T--G---GCAC-G-     1525
CPZANT        --GCC--AAA-AAAACAACAC......TCC---C---------GCA-T.....---A----GCTGT-C-     1534
A_ROD         ---TACTCCTCT-CTCAC.........GGG---C-T-C-----GT--GTTC...-TG-----G......-     1552
B_EHOA        ---TATTCCTCA-TGACA.........CCG--GA-T-------GT--ACTT...-TG-----G......-     1546
SD_MM251      --GTACACTACT-GTGGCACC......TCA---A-T-------GG--CTTT...-TG-----G......-     1600
STM_STM       ---TACACAACTAGTACC.........TCA---ACT--G----GG--CTTC...-TG-----G......-     1603
VER_AGM3      --GTATACG-GA-GTCAT.........GAC---AC---GC---TCCCGTTC...-TG-----G......-     1648
GRI_AGM677    ---TATACT-GCCCCACA..........----------G-TGCCATTC...-TG-----G......-        1585
SAB_SAB1C     ---TACACT-GCCCCGAA..........---C-------G-TGCCATTC...-TG-----G......-        1600
SYK_SYK       C-GTAT-AGT-GCCTAAT.........ACC--G--------G----CACCATTG-C---T--G......-     1522

            <- RRE                                        / RRE primary binding site
ENV_CONSENSUS TccT?GGgtTctT?ggagCaGCaGG?agcaC?ATGGGcGcagcgtCaataaCgcTgaCggtaCaG?cccg  1175
A_U455        ----T--------A--------T--A-----A--------G-----------------------G--A-     1604
B_HXB2R       ----T--------G----------A-----T----------------G---------------G--A-     1625
C_UG268A2     ----T--------G-----C-----A-----T----------A---------------------G--A-     1604
D_ELI         ----T--------G----------A-----G---------CG----G-G---------------G--A-     1616
E_TN2432      --T-T--------A----------A-----T--------G------------------------G--A-     1622
F_BZ163A      ----T--------GA-C--G-----A-----T--------G------------------------G--A-     1574
G_LBV217      ----T--------A--------T--G-----T--------G------C--G--------------GT-A-     1628
H_VI557       ----T--------G----------G-----T--------G------G-G---------------G--A-      806
O_ANT70C      --T-G---G-TC-AA-T--------T-----T----------G---C-------G--------A---A     1631
O_MVP5180     --T-G---G-GC-AA-T--------T-----T----------G---C-G-----------G-A---A     1661
O_VAU         --T-G---A-TC-AA-T--------A-----T----------G---C-G-----------G-A---A     1670
CPZGAB        ----T------C-G--T--------A-----T----------G--G---------------G----      1595
CPZANT        --T-G--TC-TC-CA-T-----T--C--T--A----------G---------C----A--A-          1604
A_ROD         --T-G--T--TC-C-C-A-------TTCTG-A--------G-----CC-G--CG--T---CT---T----  1622
B_EHOA        -TT-G--A---C-T-CGA-G-----TTCTG-A--------G-----CT-G---T--T---CT---T----  1616
SD_MM251      --T-G--T--TC-C-C-A-G-----TTCTG-A--------G-----GT-G--------C-CT---T----  1670
STM_STM       --T-G--T--TC-C-CGA-------TTCTG-A----------GC-G----------CT---T----      1673
VER_AGM3      ----A--C-----A-----T--T--G-CTG-A-----A------G-G-C-G-C--------C---T-T-A  1718
GRI_AGM677    -T--A--C-----G-----T--T--A-CTG-A----------G---C---------A--C---T-T--     1655
SAB_SAB1C     ----A--T-----G--------T--AGCTG-A-----T------G-G-C-G-C--------C---T-T-A  1670
SYK_SYK       ----G--A--AC-CTCG--T--G--C-CTG-A-----G-GC--TG-G-C-G----A--AC-C---T-T-A  1592

RRE primary binding /
ENV_CONSENSUS ?cattTa?Tg?ctGGtATagTgCAaCAgCAgaa?aAtcTgcTg???Gc?aT?gaggc?CAaCA?cA?cTg  1233
A_U455        A--A---T--T---------------------GC---------AGG--T--A-----T-----G--T---  1674
B_HXB2R       A--A---T--T-------------G--------C---T-----AGG--T--T-----G-----G--T---  1695
C_UG268A2     A--A---T--T-------------------A--C---T-----AGG--T--A-----G-----G--TA--  1674
D_ELI         A--A---A--T-------------------A--C---T-----AGG--T--A-----G-----G--T---  1686
E_TN2432      A--A---T--T-------------------R-GC---T-----AGG--T--W-----G-----G--T---  1692
F_BZ163A      A--A---T--T----A--------------GT---------AGG--T--T--A--G-----G--T---    1644
G_LBV217      A--A---T--T----C--------------A--C---T-----AGG--T--A-----G-----A--T---  1698
H_VI557       A--A---T--T-------------------A-GC---T-----CAA--T--AC----A-----G--TA--   876
O_ANT70C      CAC---GC--AAG-----------------G-C--C-----AAGA--A--AC----C--G--G--AT--   1701
O_MVP5180     CAG-G--C--AAG-----------------G-C--C-------AGA--G--AC----C--G--A--CT--  1731
O_VAU         G---C-GA-AAAG-----------------G-T--C-----AAGA--A--AC----C--G--A--CT--   1740
CPZGAB        A--G--GC--T-G----------G--A-----T------T--AAA--A--A-----T-----G--C--A   1665
CPZANT        GA----GY-CCA------T--A---------GCC---------CAA--C--A---A-A--G--A--T---  1674
A_ROD         GAC----C--G-C--G---------G--A---C-AC-G---T--GAC-TGG-CA--AGA-----AG-A---  1692
B_EHOA        GAC----C--G----G---------G------C-AC-G--CG--GAC-TGG-CA-AAGA-----AG-A---  1686
C_2238                                                    GAC-TGG-CA-AAGA-----GG-AA--    27
D_F0784                                                   GAC-TGG-CA--AGA-----AG-AT--    27
SD_MM251      GAC----T--G----G---------G--A---C-AC-G---T--GAC-TGG-CA--AGA-----AG-AT--  1740
STM_STM       GAC----T--A----G---------G--A---C-AC-G---T--GAC-TGG-CA--AGA-----AG-AT--  1743
U_SMCI2                                                   GAC-TGG-CAC-AGA-----AG-A---    27
VER_AGM3      G------C-TG----G---C----G---------G---------GCG--TG-G-----T-----G--GA--  1788
GRI_AGM677    G-----GC-TG----G---T----G---------G--CT-----GCG--TG-G--ACAG-----A--GT--  1725
SAB_SAB1C     G--G---C-CG----G---C----G---------A--------TGCG--TG-A---CAG--G--G--AA--  1740
SYK_SYK       GAC---GT--G----G---------G------C-A--GT-----GAG--AG-T-----A--G--A--C--C  1662
```

```
                                                                     RRE ->
ENV_CONSENSUS tTgcaact?aC?gTcTGGGG?at?AaacA?CTccaggC?aGagTc???GCt?T?GAgA??taccTa?agg   1289
A_U455        ---A----C--T--------C--T-----G--------A------CTG---G-G--A-GA------C---   1744
B_HXB2R       --------C--A--------C--C--G--G--------A---A--CTG---G-G--A-GA------A---   1765
C_UG268A2     --------C--G--------C--T--G--G------A-A------CTG---A-A--A-GA------C---   1744
D_ELI         --------C--G--------C--T-----G--------A---A--CTG---G-G--A-GA------A---   1756
E_TN2432      --------C--A--------C--Y--R--G--------A------YTR---G-G--A-GA------A---   1762
F_BZ163A      -----G--C--A--------C--T-----G--------A------CTG--CG-G--A-GA------C---   1714
G_LBV217      --------C--A--------C--T-----G--------A------CTG---G-A--A-GA------C---   1768
H_VI557       ---A-G--C--G                                                              888
O_ANT70C      C--AGG--AT-TR-A-----T--C-G---A----GA--TC-CC-GCTA--CT-A--A-CC-TA---C--A   1771
O_MVP5180     C--AGGT-AT-T--A-----T--T-G---A----GA--TC-CC-GCAA--CT-A--A-CCCTTA--C--A   1801
O_VAU         C--AGG-CAT-T--A-----T--T-G---A----GA--TC-CC-GCTA--CT-A--A-CC-TTA--C--A   1810
CPZGAB        --------AT-AA-------AG-A-----A--A--A--C---T-GCTT---G-A--A-GG-----GC---   1735
CPZANT        C-A--G--CT-G--------AG-A-----A--------A---A-GCTT--AG-C----AG------AGA-   1744
A_ROD         ----G---G--C--------A-CG---A-C--------A------ACT---A-A----AG------C---   1762
B_EHOA        ----GG--G--C--------G-CG---A-C--------A------ACT--CA-C----AA--T--CA-A-   1756
C_2238        ----GG--G--C--------A-CT---A-C------A-T------ACT--CA-C----AG------A---     97
D_F0784       ----G---G--C--------G-CT---A-C------A-T------ACT---A-C----AG------A---     97
SD_MM251      ----G---G--C--------A-CA--GA-C------A-T--G---ACT--CA-C----AG---T--A---   1810
STM_STM       ----G---G--C--------A-CC--GA-C------A-T------ACT---A-C----AA------A---   1813
U_SMCI2       ----G---G--C--------A-CT---A-C------A-T------ACT--CA-C----AA------A---     97
VER_AGM3      ---A-G--G--CA-T-----TG-G---A-C---A-T--CC-C---ACA---C-T----AG------G---   1858
GRI_AGM677    ---A-G--G--CA-T-----TG-G---A-C---A-T--CC-C---ACA---C-C----AG------G---   1795
SAB_SAB1C     C-AA-G--C--TA-T-----TG-A---A-C---A-T--CC-C---ACT---C-T----AA------G---   1810
SYK_SYK       --AGGCT-A--T--------TG-C---A-C--TA-C--CC-GC--ACA---T-A----CG--T--CAGA-   1732

ENV_CONSENSUS AtCAg?agct?cTaaa??ttTGGGG?TGcgc??g?Aaaca?gtcTGccacACtactGTgcc?TGG?????   1344
A_U455        ----AC----C---GGAA-------C---T-TG-A--G-TCA-----AC------------C---.....   1809
B_HXB2R       ----AC----C--GGGGA-------T---T-TG-A----TCA-T---AC----G-------T---.....   1830
C_UG268A2     ----AC----C---GGGA-------C---T-TG-A----TCA-----AC----G-------T---.....   1809
D_ELI         ----AC----C---GGAA-------T---T-TG-A-----CA-T---AC-----A------C---.....   1821
E_TN2432      ----AA----C--RGGAC-------Y---T-TG-A---ATCA-----AC----G-------C---.....   1827
F_BZ163A      ----AC----C---GGAC-------C---T-TG-A----TCA-----AC-----A------C---.....   1779
G_LBV217      ----AC----C---GGGA-------C---T-TG-A----TCA-----AC-----A------T---.....   1833
O_ANT70C      -----C-A--C----GCC-A-----C--TAAAG-A--G-TA------T----AT-A--AAAA---.....   1836
O_MVP5180     -----C-A-GC-----CC-A-----C--TAAAG-A----TAA----TT----AT-A--AAAA---.....   1866
O_VAU         -----C-A--C--T--CC-G-----C---AAGAAT-G--TAA-----T----AT-A--AAAG---.....   1875
CPZGAB        -----C--A-T--GGGCC-G-----C---T-TG-A--GGCT-----TT-T--C--G-----T---.....   1800
CPZANT        ----AC-A--AT-G-GCC-C-----T--T--TGAC--GGTGAC---T--------G-----T---.....   1809
A_ROD         -C---GC--GG-----TTCA-----A--T--GTTT-G---A-----------------A--A---GTT..   1830
B_EHOA        -C---GCA-AA-----TTCA-----A--T--TTTC-G---G-----T-----G-----A--A---GTA..   1824
C_2238        -C--AGCAAAA-----TTCA-----A--T--CTTT-G---G--T-----------------A---GTG..    165
D_F0784       -C---GC--AG-----TTCA-----A-----GTTC-G---G-----------C-----AGAA---CCA..    165
SD_MM251      -C---GC--AG-----TGC------A--T--GTTT-G---A-----------------A--A---CCA..   1878
STM_STM       -----GCA-AG-----TTC------A--T--GTTT-G---G-----------C-----A--A---CCA..   1881
U_SMCI2       -----GCA-GG-----TGC------A--T--ATTT-G---A--------------G--A--T---GAA..    165
VER_AGM3      -C---GC--GGT-G--TGC------G-----AT-G--G--A-----T--T--A--C--A--G---CAGTG   1928
GRI_AGM677    -----GCA-GG-----TTCA-----A--T--GT-G-----A--A--T-----C--A-----A---AAGTA   1865
SAB_SAB1C     -----GCA-GGT-G--CA-------A--T--TTTC-GG--G--G--T-----G--G--C-TT---AAGTA   1880
SYK_SYK       -C---GCCA-CT-GTCAAA------A-----ATTC--G--GA-------T--AG-A---A-G---GAG..   1800

ENV_CONSENSUS ?AAt???acttgg?????????aa????????t??a????????????????????????????c??att   1361
A_U455        .--CTCT-G----......AGT--TAAATCTCAGG-G...........................GAC--A   1845
B_HXB2R       .---GCT-G----......AGT--TAAATCTC-GG-A...........................-AG---   1866
C_UG268A2     .--CTCT-G----......AGT--TAAATCTC-AGGG...........................GAT---   1845
D_ELI         .--CTCT-G----......AGT--TAGATCTC-AA-T...........................GAG---   1857
E_TN2432      .--CTCC------......AGT--TAGATCTT-TG-A...........................GAG---   1863
F_BZ163A      .--CTCT-G----......AGT--TAAATCTCAGG-G...........................GAG---   1815
G_LBV217      .--CACT-G----......AGC--TAAATCTT-TA-T...........................GAG--A   1869
O_ANT70C      .---AGA--A---ATA...GGA--CGAAAGC....................................---   1866
O_MVP5180     .--CACAT-A---......TCAGGAAGATATAATG-TGAC........................AGT---   1905
O_VAU         .---AAA--A---GGAGGAGAT--TGAATCA....................................---   1908
CPZGAB        .--CAACT-----CCTGGGAGC--TTCCACA...G-T...........................GAC---   1839
CPZANT        .---AATT-C---......GTA--CTTCACGCAAACATGTGCAAAGAACAGCAGTGATATACAATGT---   1872
A_ROD         .---GATT-C.....................T-AGCA...........................-CTGAC   1851
B_EHOA        .---GAAT-C.....................C-TA-G...........................-CAGAC   1845
C_2238        .--CGAC-G-.....................A-GC-G...........................-CAGAG    186
D_F0784       .---AGT---.....................C-CACA...........................-CTGAC    186
SD_MM251      .---GCA-G-.....................C-AACA...........................-CAGAC   1899
STM_STM       .---GATT--.....................T-GGTA...........................-CGGA-   1902
U_SMCI2       .---AAC---.....................T-GC-A...........................-CTGA-    186
VER_AGM3      G---AAT-GG........................ACC...........................-CTGA-   1947
GRI_AGM677    T---AAC---......................................................-CT-AG   1881
SAB_SAB1C     T---AAC---......................................................-CCGAC   1896
SYK_SYK       .--AGCATG-G-AAACAACTCC--CTTCTGTCCCA-A...........................-CACAG   1842
```

```
ENV_CONSENSUS TGG?a?aacatgAC?TGGcaggaaTGGgA?aga?aaaTt?acaattt?a?ag??aa?aTatat?a??ta?  1416
A_U455        ---A-T--------C----T-C-------G-A-G-----AGT-G--AC-C--GC-TA------C-AC--A  1915
B_HXB2R       ---A-TC---C---C---AT---G-----C---G-----A------AC-C-AGCTTA---C-CTCCT--A  1936
C_UG268A2     ---G-T--------C---AT-C-G-----T---G-----AGT----AT-C--GA-CA-----CAGGT-GC  1915
D_ELI         ---C-G--------C---AT---G-----A---G-----G------AC-C--GCTTA------AGCT--A  1927
E_TN2432      ---A-C--------M---ATA--------R---G-----AG-----AC-C-AACC-A------G-GA--C  1933
F_BZ163A      ---G-G--------C---AT-------C-A-A-G-G---AG-----ACTC-AATG-AG----CAGGT--A  1885
G_LBV217      ---G-C--------C---ATA--------A--GG-----A----C-AC-C-CACC-A-----CTCCC-GC  1939
O_ANT70C      ---G-C-C-T-A--A-------------TC-GC-G--AAG---CA-A-GCTCC-CC------G-GGA-A  1936
O_MVP5180     ---G-C---C-T--A------C-------CCA-C-C--AA-----G-A-GCTCC-TT------G-TGA-A  1975
O_VAU         ---G-TG-GT-A--A------C-G-----TCA-C-G--AA----CG-A-GCTCCTTC------G-AAA-A  1978
CPZGAB        ---GGG--TC-A--A------C-------T-A-TT-G-GTCT--C-AC-C--GG--A--T-T-GGTC-GT  1909
CPZANT        ---G-A--T-----A-----A--------C---TT-G-AC-G----CA-C--GAC-G------A-TA-CT  1942
A_ROD         ---G-C--T-----G-------------A-A-C--G-CCG-T-CC-GGAG-CA--T--CAG-A-AAGTT  1921
B_EHOA        ---A-T--------A-----AC-------G--GC--G-CCG-TTC--GGAT-CA--T---ACAA-AT--C  1915
C_2238        ---C-A--------T-----AC-------A-A-C-G---GC-TTC--GGAG-AT--T--CACAG-GC-GT   256
D_F0784       ---A-C--T-----T--------G-----A---C-GG--G-TTTCC-AGAG-CA--T---ACAC-AT--T   256
SD_MM251      ---A-C--TGAT--T-----A--G-----GC--A-GG--G--TTC--GGAG-AA--T---ACAGCCC-CC  1969
STM_STM       ---A-C--T-----A-------------G---A-GG--G--TT-C-TGAG-CA--T---ACAC-AC-GT  1972
U_SMCI2       ---A-T--T-----T-----A--------A-T-A-G--CAGAG--C-GGAG-CA--T--C-CAG-ATC-T   256
VER_AGM3      ---A-T--T-----T----T---------A---C-G--ATCGT----GGA--GT--C---ACAACACA-T  2017
GRI_AGM677    ---G-C--T-----T---TT---G-----G---C-----A-TGCC--GGA--GC--C---AC-C-AC--T  1951
SAB_SAB1C     ---G-A--------G-----A--------G--GC----AG-A--G-ATGA--CT--C--CAG-AGGA-TT  1966
SYK_SYK       ---A-A--T-----A-----CAG------ACA-G-GG-AG-------G-CT-ACC-T--TG--GGAT-GC  1912

ENV_CONSENSUS TagaagaagCacaaa??CAgcAggaaaagAAtgtg?a?ga?ttac?ggaatTagataaaTGGgc???ttT  1477
A_U455        -T--G---T-G--G-AC---------------AACTA--C---TT--C---G--C--G-----AAA-C-  1985
B_HXB2R       -T------T-G----AC-----A----------AAC-A--A---TT-----------------AAG---  2006
C_UG268A2     -T-----CT------AC--------G--A----AAA-A--T----TA-C---G--C------CAAAA-C-  1985
D_ELI         -T--G---T-G--G-CC-----A----------AAA-A--A--GTT------G--C--G-----AAG---  1997
E_TN2432      -TAC----T-G--G-AC--------C-G-----AAA-G--T--GTTA-----G-----------AAGCC-  2003
F_BZ163A      -T---A--T----G-AC---------------AAC-A-GA---TT--C---G--C--------AAG-C-  1955
G_LBV217      -C------T-G--G-TC---------------AAC-A--C---TT--C---G--C--------AAAGC--  2009
O_ANT70C      --C--A-G------GTA---------C-A----A-A-AA-G--G-T---G------G-------CTC-A-  2006
O_MVP5180     --C---C-------GAC--A--------------AA-A-CA--GTT---GC-----G-------CTC-C-  2045
O_VAU         --C----G------GAA--A-----G--A----A-A-A--A--G-T---G------G-------CTC-A-  2048
CPZGAB        ----G---------TCA--A--A--G--A----AAAGG--CC--TT------G---C-------CAGC--  1979
CPZANT        --C--AT------TGAG--A--A--G-GA---AAAA-G--A---TAT---C----C------AGCTCA--  2012
A_ROD         -----C-G-------TT-----A--G--A---A--T-T--AC---AAA-----A---GC----ATAT---  1991
B_EHOA        -------G-----G-TA--A--A--------CA--T-T--G----A-A-----A--C------ATAT---  1985
C_2238        ----GC----T----TA--------G--A---A--T-T--G--G-A-A-----A---GC----ATGT---   326
D_F0784       ----G----------TT--------G--A--CA--T-T--G----AAA--C--A----C----ATATA--   326
SD_MM251      -------G-------TT--A--A--G-----CA--T-T--A----AAA-G--GA---GC----ATGTG--  2039
STM_STM       ----------T---GTT--A-----------A--T-T--G--G-AAA--C--A---GC----ATGTA--  2042
U_SMCI2       -G--G--G-------TA--A--A---------A--T-T--G--G-AAA-----A---GC----ATGTG--   326
VER_AGM3      ----G-----CAG-GCA---G----G------T--G-T-CA-AC-AAA-----AG--GT---T-AGA---  2087
GRI_AGM677    -G-------------AT---G-ATC-------C--G-TCTG-AC-A-A----G---G-T---T-AGGG--  2021
SAB_SAB1C     ----GC-------TGAA---G-AC--------T--G-TAGT-AT-AAA----G-TGTCT---T-AGAC--  2036
SYK_SYK       -CAG----------GAA--A-----G-GA------C-T--TC--ACAA-G---C-GG------ACTCG--  1982

ENV_CONSENSUS ?tGgaattGGtTtgacaTaaCaaa?TGG?TgtggtataTaaaaatag??aTaat?atagTaggagg??Ta  1539
A_U455        G--A---------A----CT----T---C----------T-G-C--TTTG----A----------CT--  2055
B_HXB2R       G------------A----------T---C----------T--TTC-----G----------CT-G     2076
C_UG268A2     G----G-------------------A---C----------TTC-----G----------CT--      2055
D_ELI         G------------AG--------C-A---C----------TTC-----G---A-------CT-G     2067
E_TN2432      G----G----------------T---C----------TTT-----G----------TT-G         2073
F_BZ163A      G------------AG----T----C---C----------TTC-----G----------CT-G       2025
G_LBV217      G----G-------AG-------C-A---C-A----------TTT-----G----------TT--      2079
O_ANT70C      T---------C----------T--A---T----------CA-----C----------CAC--       2076
O_MVP5180     T--------------------T--A---T----------CT-----C-----G----CAC--       2115
O_VAU         T---------C----------T--A---T----------CT-----C----------CAC--       2118
CPZGAB        G----------------------A---C----------TTTCC-T--GGC---------AA-C     2049
CPZANT        A----------------------C-A---C-A----------TTT--T--G----------CTA-T    2082
A_ROD         TG-C------------T----CTCC---G-CAA------TC--TAT-GAG-GC-T---A---T--CAG--  2061
B_EHOA        CA-T------------T-C--CTCC---A--GCA--C--C-GGT---GAT--TAT------AT---AA--  2055
C_2238        TG-C------C----TC-C--CTCT---G-AAAA-----CT-TT---GAT-TTAT-------C---AG--   396
D_F0784       TG-C--C---------C-T--TTCT---A-TAA-------C-GTAT-GAG-CT-C-------T---TA--   396
SD_MM251      TG-C------------C-TG-TTCT---A-AAA-------C--TAT-GA--TTATG----T-T---AG--  2109
STM_STM       TG-C------------C-T--CTCT---G-AA-A--C---C--TAT-GAG--TATT-----TAT---AT--  2112
U_SMCI2       TG-C--CC--------TT-T--CTCT---G-CAAA-----CT-T----GAT-TTATG------T---AA-T   396
VER_AGM3      C---TC------C--TT-CT----G---C--AACAT-C---------GAT-TT-GGAT---CT---TA-T  2157
GRI_AGM677    C---TCA-----CTCAC-GT---CT---T-AG-C---G-T-------GAT-TT-AG-GA---TGATTA-T  2091
SAB_SAB1C     T---TCG----------C-C-----A---T-TG-A-GG--G--G----CC--C--GG-TA---C---CA-C  2106
SYK_SYK       G----GC---------T-GT-G--A---T-C-TT---C----G----GAT-TTATG--A------CTT--  2052
```

```
                                                                3' sj \/  \/ 3' sj
ENV_CONSENSUS aTagg?ttaAgaaTagtt?t???tgta?t???tata?Taa?tAg??TtaggcagGgaTAt?caCCt???c   1592
A_U455        -----A----------T-TAC---GC-TTC----A-C-A---AG-------------CT-----...-   2122
B_HXB2R       G----T----------T-TGC----C-TTC----G-G-A---AG-------------T----A...T   2143
C_UG268A2     -----T---------A--T-TGC---GC-CTC----G-G-A---AG-----A-------CT----A...T   2122
D_ELI         -----T----------T-TGC---GC-TTC-T--G---A---AG-------------CT-----...-   2134
E_TN2432      -----T---------A--T-TGC---KC-TTC----G-G-A---AG-------------CT-----...T   2140
F_BZ163A      -----C----------T-TGC---GC-TTC----G---A---AG-----A---C---CT-----...T   2092
G_LBV217      -----T----A------T-TGC---GC-TTC----G-G-A---AG----------CT-----...-   2146
O_ANT70C      G----GG-G---G-TA-CA-GATA---C-TAA----G-G-AA-ACA--------------CA---C...-   2143
O_MVP5180     -----TA-----G-TA--A-GATAA--C-TAA-C--G-G-AG-ACA--------------CA---C...-   2182
O_VAU         -----TG-----G-----A-GATA---C-TAA-C--G---AG-ACA--------------CA---C...-   2185
CPZGAB        -----G--------TA--A-GACA--TT-CTCAG--G-C-GG--AG----------C--CT-----...-   2116
CPZANT        G----AC-------TT-GC-TGTAT--G-TAG-TGCT---GA-AGG--------------CAT---...-   2149
A_ROD         ----CT----------GA-ATA----G-ACAA--GT---G---GC----AA----C---AGG---GTTT   2131
B_EHOA        G---TA----------CAA-ATACA-TA-ACAG--GC--GCA--GC-----A----C---AGG--AGTAT   2125
C_2238        ----TG----------C-A-CTA----G-ACAA--GT---TG--AC-----A----C                453
D_F0784       ---CTG---------A-AA-CTA----G-ACAA--GT--GC--AGT-A--TA--                   450
SD_MM251      ---CTG----------GA-CTA-A--G-ACAA--GC--GC--AGT-A--------G---AGG--AGTGT   2179
STM_STM       G--ATG------G---CAA-CTA-A--A-GCAAT-GT--GC---GC-----A-A--T---CGG---GTGT   2182
U_SMCI2       -----G---------A-AG-CTACT--A-ACAAT--T--GG--AGC-----A--                   450
VER_AGM3      -----A------T-GC--TATACA---TATTC-TGCA--GC---GG----------T--CT-T---...-   2224
GRI_AGM677    C----A------T-T-CATGGGTAT--TGGGGATGTA-C-GA-ATA--------------AAT---...-   2158
SAB_SAB1C     ----TAGCT---G-TC-GC-AGTAA--A-AGGA--TC-T-GG-AGT----AA-A--G---G-----...-   2173
SYK_SYK       G-GCTT-----GC-----AGTTT-AGTG-AGGA--CA---AG-ATT-GTT-GGA--C---GTT---ATTT   2122

ENV_CONSENSUS Tctc?tt???????????????????cagAccC?tatCCaaaa?cag??gGaa???ccagaca?gccaGa   1634
A_U455        -G--G--T.....................-------T-GC--C--TC-CAGA--GT...-TC-G--G-----G   2171
B_HXB2R       -A--G--T.....................-------ACC---C--CC-C-AG--G-...--C----G---C--   2192
C_UG268A2     -A--G--T.....................-------T--C--C--CC-C-AG--G-...--C----G---C-G   2171
D_ELI         -G--G--T.....................-------TCC---C-GCC-C-AG--G-...--C----G---C--   2183
E_TN2432      -G--TC-C.....................-------C--C---TC-T---AG----...-TC----GA--C--   2189
F_BZ163A      -G--A--T.....................-------A-----C--GC-C-AG----...--C----G---C--   2141
G_LBV217      -A--A--C.....................-------T-GC---CC-T---AG----...--C----G---T--   2195
O_ANT70C      ----G--A.....................----T--CC-A---TC-C--AGA----...G---GA-C-----G   2192
O_MVP5180     ----G--G.....................----T--C-G---C-C-C-G-CA----...G----A-C-----G   2231
O_VAU         ----G--A.....................----T--CC------C-A--AGC----...GT--GA-C-----G   2234
CPZGAB        ----G--G.....................-------T-----C-GTC---AG----...-A--GA-G--TC-G   2165
CPZANT        -G--A--T.....................----T--C--C------C---CA---T...-----GCA------   2198
A_ROD         ----T-CCCCCCCCGGTTATATCCAA----T--A------C--GG-CCG--G-CAG----C--ACGA---   2201
B_EHOA        ----C-CCCCTCCCTCTTATACTCAA----T--C-----GC--GG-CCG--G-CAG----C--ACGA---   2195
SD_MM251      ----T-CCCCACCCTCTTATTTCCAGT----T-A--C----C-GG-CCC--C-CTG---AC--GAGA--G   2249
STM_STM       ----T-CCCCTCCCTCTTGTCGTCAG----T--C------C--GGGCCA----CAG---AC--AAGA--G   2252
VER_AGM3      -T--TCCA.....................----T--A------CCCGTG-AA--G-CAG-------ACG----   2276
GRI_AGM677    --C-C.....................----T--A------C-GTTCAGC----CGG-------ACGG--G   2207
SAB_SAB1C     ----C-CCCTCCCCTCTTCTCATTAA                                              2199
SYK_SYK       -GCAGAAC.................-CT----AGGG--GG--GG-CCCA-GCAAA--C-CAGACGA---   2174

ENV_CONSENSUS a??aa?agaagaag?aggtggaga??aaggcagagacAgatcc????????????tgc??ttg????tg   1679
A_U455        -AG--TC-------A---------GC------A---------G.....A-T-GC--A...G--   2226
B_HXB2R       -GG--T--------A---------GAG--A-------------.....A-T-GA--A...G--   2247
C_UG268A2     -GA--TC-------A---------GC---A-------------.....G--AGA--A...A-C   2226
D_ELI         -GG--C--------A---------GCG-----------------.....G--AGA---...C--   2238
E_TN2432      -AG--TC-------G------C--AC----------A------.....G---GC--A...G--   2244
F_BZ163A      -GA--TC-------G---------GC------------------.....G--AGA--A...G-A   2196
G_LBV217      -AG--TC-------G------C--GC---A-------------.....G---GC--A...G--   2250
O_ANT70C      -AG--C--G--G--G---------AG-------GCC---G-GG.....A-A-CC-C-...CC-   2247
O_MVP5180     -AG--C--G-----A---------AGG--A---GCC--AG-GG.....ACAGCC---...CCA   2286
O_VAU         -GG--C--G-----G---------CG---A---GCG---G-GG.....ACT-CA---...CC-   2289
CPZGAB        -GA--TC--C----G------C--GC---A----AG------A.....G--AGA--A...G--   2220
CPZANT        -GA--T-AG-----A-------AGAA---A---GAT---G-GG.....AG-GCC---...CA-   2253
A_ROD         -...-C--------AC------AGCA-C--TG---------A-.....TG--CC-G-CCGA-A   2256
B_EHOA        -...-C--------G-------A-CA-C-AGG-CT-------T.....TG--CC-G-CAGA-C   2250
SD_MM251      C...-A-----G--AC--------AGGC--TG-CA----C---.....TG--CT-G-CAGA-A   2304
STM_STM       -...-C--------G---------CAG---TG--AT--AC---.....TG--CT-G-CAAA-A   2307
VER_AGM3      -...GGGCC--G--A---------CA-GC---AGA----C---.....GA--CT-G-CAGAAA   2331
GRI_AGM677    -...............--GCA---CAG---TG----A--CAG-AGCAGCAAATTGA-AAGA---...CA-   2259
SYK_SYK       -.................--AGGCA-TG--------GAA.....GGA-TCAAC...G-C   2208
```

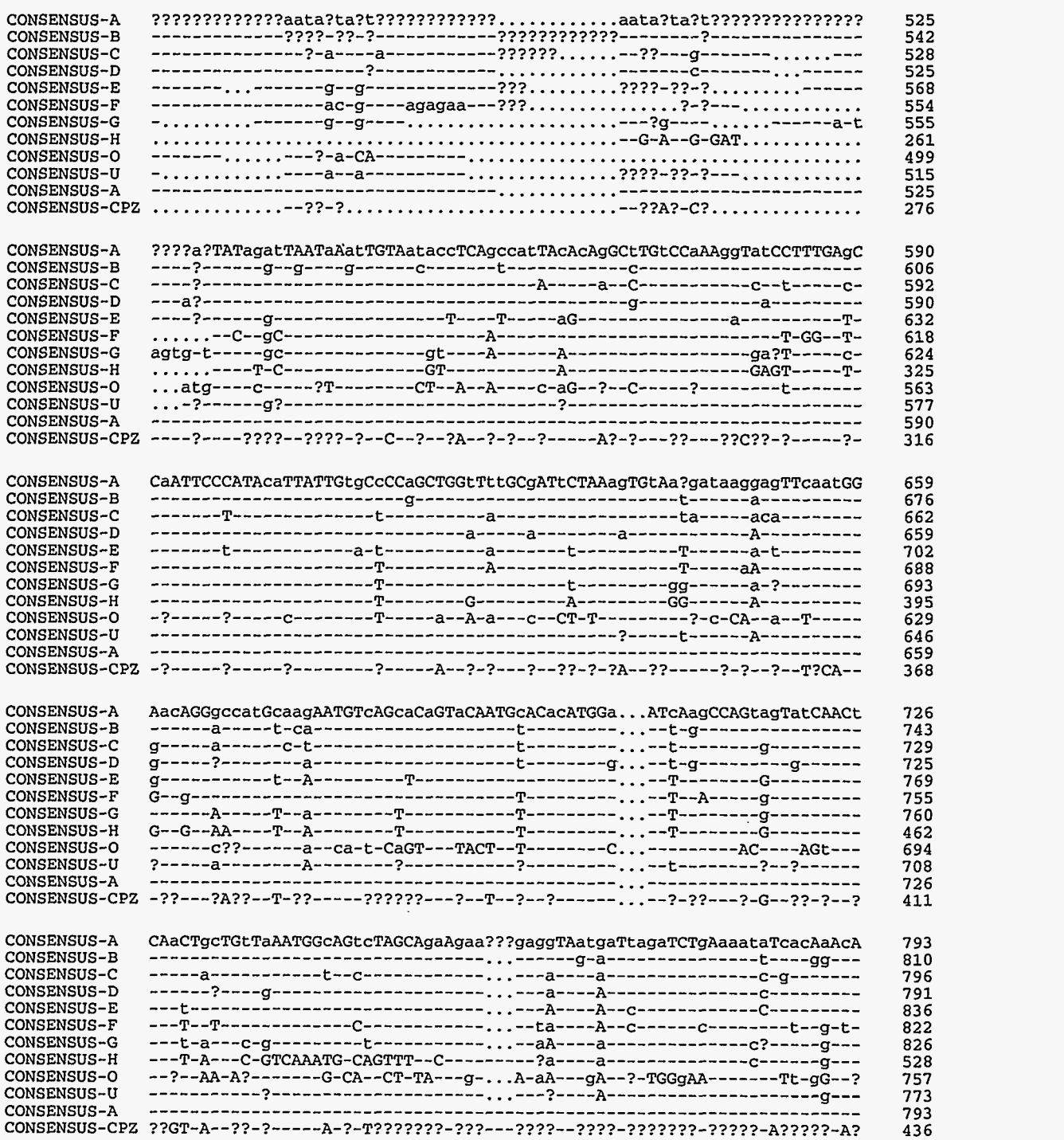
```
ENV_CONSENSUS .?aa???gg??????????????????????????????????????????????????ttctt??cac  1691
A_U455        AGC...--A.....................................................-----AG-GA  2242
B_HXB2R       A-C...--A.....................................................-C---GG---  2263
C_UG268A2     AGC...--A.....................................................-----AG---  2242
D_ELI         A-C...--A.....................................................----CGG---  2254
E_TN2432      AGC...--A.....................................................-----AG---  2260
F_BZ163A      G-C...--A.....................................................-----AG-T-  2212
G_LBV217      AGC...--A.....................................................-----GT-GA  2266
O_ANT70C      C--...--G.....................................................-----GC---  2263
O_MVP5180     CC-...--A.....................................................-----GCA--  2302
O_VAU         C--...--G.....................................................-----GCAT-  2305
CPZGAB        G--...--G.....................................................-G---GC---  2236
CPZANT        C-C...--G.....................................................-----CG---  2269
A_ROD         GC-TATATA.....................................................CAT--CCTGA  2275
B_EHOA        G--TACATC.....................................................CA---CC-GA  2269
SD_MM251      G--TATATT.....................................................CAT--CCTGA  2323
STM_STM       G--TATACT.....................................................CAT--CCTGA  2326
VER_AGM3      G--TCT--CACAGCAGAGTGGAAGAGCAACTGGTGCAAGCGATTGACCAATTGGTGCTCGA--AGCAGCA  2401
GRI_AGM677    G--...-AG.....................................................-C--CAA---  2275
SYK_SYK       TCT...ACC.....................................................---AGCAG-G  2224

ENV_CONSENSUS tt???tg??acga?ct?cgga?cctg??c?t?tg?a??taccaccgcttgagagactta?ta????????  1737
A_U455        --GCC--GG----C--G----A----TG-C-C-TC-GC--------A----------CGCC........  2304
B_HXB2R       --ATC--GG----T--G----G----TG-C-C-TC-GC----------------------C-C.....  2325
C_UG268A2     --GCC--GG----T--G----G----TG-C-C-TC-GC----G----------------C--.....  2304
D_ELI         --ATC--GG----C--G----G----TG-C-C-TC-GC----------------------A-C.....  2316
E_TN2432      --GCC--GG----T--A----G----TG-C-T-TC-GC--------------------CA-C.....  2322
F_BZ163A      --GTC--GG----C--GA---A----TG-C-C-TC-GC-------A-C----------C--.....  2274
G_LBV217      --GCC--GG----CT-GA---G----TG-C-C-TC-GC-------AA----------CA-C.....  2328
O_ANT70C      -GTTG-ACACG--C--CA-A-CAA-AATAT-G--G-CT-------T------CA-----GC-.....  2325
O_MVP5180     AGTTG-ACACG--T--CA---CAA-AAT-T-G--G-CT-------T------CA-----A--.....  2364
O_VAU         -GTTG-ACACG--C--CA---CAA-AAT-T-G--G-TT-------T------CA-----GCC.....  2367
CPZGAB        --ATC--GG----C--C--A-A----GGGA-T--G-GC-----GA------C-AG----GCC.....  2298
CPZANT        -CTTG                                                               2274
A_ROD         -C...C-CC-GCTGA-T--CCT-T--AC-...A-ACTA---AG-AT--GC--G------C--.....  2331
B_EHOA        --......CG-C-A--GA--GA---CTTGA-T--GCTA---AG-G---GC---AC----T--.....  2325
SD_MM251      -C...C-CC-ACTGA-A--CCT-T--ACT...--GCTA-T-AG-AA--GC---AC---GC--.....  2379
STM_STM       -C...C-CC-GCTAG-A--CCT-T--ACT...--GCTG---A--AA---C--G-CT-GTC--.....  2382
VER_AGM3      -C.............................--GCTA---A--A-T-GCCTGAC-C-CC--GTTCATCT  2441
GRI_AGM677    C-.....................................TCGA-GA-C-ACA---GGTGG.........  2298
SYK_SYK       AA.....................................................................  2226

ENV_CONSENSUS ????t?????at?????????????????gag??t?????aa?t?ct?ggac?????a???????????  1754
A_U455        ....-TG...--TGTA...........GC---AGCTGTGG--C-T--G----GCAGC-GCCTCAAGGGA  2355
B_HXB2R       ....-TG...--TGTA...........AC---GA-TGTGG--C-T--G----GC..............  2361
C_UG268A2     ....-TG...--TGCA...........GC---AGCAGTGG--C-T--G----GCAGC-GTCTCAGGGGA  2355
D_ELI         ....-TA...--TGCA...........GT---GA-TGTAG--C-T--G----GC..............  2352
E_TN2432      ....-CG...--TGCA...........GC---GGCTGTGG--C-T--G----ACAGC-GTCTCAAGGGA  2373
F_BZ163A      ....-TA....-AACA...........GC---AA-TGTGA-CAGG.....................GGG  2304
G_LBV217      ....-TG...--TGTA...........AC---GACAGTGG--C-T--G----GCAAC-GCCTCAAGGGA  2379
O_ANT70C      ....-CAGGG--CCAGAAG.......GT--TCAGCTATCTGAGG--T----TGTGG-TC......CTA  2376
O_MVP5180     ....-CAGGG--CCGGAGG.......CT--TCGACTACCTGGGA--G----TGTGG-TC......CTG  2415
O_VAU         ....-CAGAG--CCAGAAG.......TT--TCAGACACCTGGGA--T----TATGG-TC......ATA  2418
CPZGAB        ....-GC...-ATGTG...........TG---ACAACTCA-GACGT-G----ATCTA-TACTTCACAGC  2349
A_ROD         ....-CCAGG-GCTTC...........CT--CCC-CCAACTCA-CTACCAGAAT..............  2370
B_EHOA        ....-CAAAG-CCTTC...........CAA-CCC-CCAACC-G-A--CCA--CG..............  2364
SD_MM251      ....-CGAGAGCATAC...........CA--TCC-CCAACC-A-A--CCAGAGGCTCTCTGCG...ACC  2430
STM_STM       ....-CCAGG--CTAC...........CA--CCC-CCACCC-ACCT-CCA-AGGATCTCCAGG...ATA  2433
VER_AGM3      TAGGAGC...GCTTTC...........CA-TACA-ACAAT-TGGG--T--GGAACTC-AAGCC...GCA  2493
GRI_AGM677    ....CTC...-ACTTCAAGAGCTGCAGCTT---AA-AAGGACTTGGTGTTACAACATCTGCCTGACCCTC  2361
SYK_SYK       .....................................................................  2226
```

```
ENV_CONSENSUS ct????aggg???gggaag?cctcaga??t?tg?g???t?t??tgcaatat?????????????????? 1787
A_U455        --GAGACT--GGT----G-G-----A-TA-C--T-GAA-C-TC---TG---..................  2406
B_HXB2R       .....----GGT------C-----A-TA-TG-T-GAA-C-CC-A--G---..................  2406
C_UG268A2     --ACAG----GGT------C---T-AGTA-C--G-AAGCC-TG----G---..................  2406
D_ELI         .....----GGT----CAT-----A-TA-C--T-GAA-C-CC-A--G---..................  2397
E_TN2432      --AAGAC---GGT------G-----A-TA-C--G-GAA-C-TC--TT----..................  2424
F_BZ163A      --GAGG----GGT------C-----A-CT-C--G-GAA-C-CAC--TG---..................  2355
G_LBV217      --GAGACT--GGT----G-G-----A-TACT--T-GAA-C-CC---TG---..................  2430
O_ANT70C      GGGCAG-A-ATAATTA-T-TTTG----AT-TGTGCAGC-G-AACA-----C..................  2427
O_MVP5180     GGACAA-A-ACAATT----CTTGT---CT-TGTG-AGC-G-AA--------..................  2466
O_VAU         GGGCAG---ACAATT----CTTG----CTCT-TAAAGC-A-AA-A-----C..................  2469
CPZGAB        --CAGACT-CTAC----GAGA--GT-CCTGC--G-GGGAA-TA-T------..................  2400
A_ROD         ...CTC--A-ACT--CTGAGA--T---ACA......GCCT-CT--------GGGTGC.............  2418
B_EHOA        ..................-.-----GCT-CCACCTGCATACC-C-G----GGGATC.............  2400
SD_MM251      --ACGA----TTC-A----T------GAC-GAACTGACCTACC-A------GGG.................  2484
STM_STM       --GCAG--AATTA-A----T-G-----CTAGGAGCAGCCTATT-A--G---GGGTGC.............  2490
VER_AGM3      GCACAA.........----CAG-TGTCGC-T--GCACGCC-TGCA---A-C...................  2535
GRI_AGM677    --CATATTCATCA--AC--CAG-GG--TAC.............C----G---..................  2400
SYK_SYK       .....................................TC-T-GAG-----C-TTGGAAGCAGGACAGCAGC  2260

ENV_CONSENSUS ??????????????????????????????????????????????????????tgG?g?????????  1791
A_U455        ..........................................................---G-T........  2412
B_HXB2R       ..........................................................---A-T........  2412
C_UG268A2     ..........................................................---G-T........  2412
D_ELI         ..........................................................---A-T........  2403
E_TN2432      ..........................................................---G-C........  2430
F_BZ163A      ..........................................................---G-T........  2361
G_LBV217      ..........................................................---G-T........  2436
O_ANT70C      ..........................................................---CTA........  2433
O_MVP5180     ..........................................................---CTA........  2472
O_VAU         ..........................................................---CTA........  2475
CPZGAB        ..........................................................---G-A........  2406
A_ROD         ..........................................................GAG---ATC........  2427
B_EHOA        ..........................................................AGC---TTC........  2409
SD_MM251      ..........................................................---A-CTAT.....  2493
STM_STM       ..........................................................ATC---ATC........  2499
VER_AGM3      ..........................................................GC-G-CTAC.....  2544
GRI_AGM677    ..........................................................G--CTC........  2406
SYK_SYK       TTTGGAGGACAGTCTGCAGCAGCTTCAGATCTCTGATTAGACAGCTTACTATCACC---G-ATTCATCAG  2330

                                                <- rev cds end
ENV_CONSENSUS ??????????caggaact?aaaa?tagtgcta??ag??tgcttga?gc??tagca?t?gcagtagc?ga?  1838
A_U455        ..........-G------A----T--------TT-CTT-------C--TG-----G-A--------T-GG  2472
B_HXB2R       ..........--------A--G-A-------GTT--CT----CA-T--CAC---CA-A--------T--G  2472
C_UG268A2     ..........-A-----A----AG-------TT--TT-------TA-CA-----A-A-------T-T--A  2472
D_ELI         ..........--------G-GG-AC-------GT--CT--T----T--CA-----A-A--------T--G  2463
E_TN2432      ..........--------A----T--------TTTCTT-----A-T--TAC----A-A--------G-GG  2490
F_BZ163A      ..........--------A--G-A--------TTG-CT--T--A-CA-CAC----A-A-T------T--G  2421
G_LBV217      ..........--------A--G-A--------TT-ATT-------TA-AG-----A-A---AC---TA-C  2496
O_ANT70C      ..........--A---T-GC-G-A--------CA--CT----A--CA-AC-T---G-G--------CA-T  2493
O_MVP5180     ..........--A---T-G----A--------CA-ACC-------TA-TA-T---G-GT----T--CA-T  2532
O_VAU         ..........--A---T-GC---C--------CA-ATC-A--A--TA-TG-T---G-G-----T--TA-T  2535
CPZGAB        ..........A----G--A----T--------TA--CC-------T--AAC---TA-T--------A--A  2466
A_ROD         ..........--A---...GC-TTCCAG--CGCCGCGAG-GC-ACAAGAGAGA-TC-T--G-GC--GTGC  2484
B_EHOA        ..........--A---...GC--TCCAA--AGCAGCCAG-GC--CG-GAGAGA-TC-T--GAGC--G-CG  2466
SD_MM251      .......TTC--T--G...GCGGTCCAA--CGGCT-GAGATC--CGA-AGA-A-TC-T--G-GC--GTGG  2553
STM_STM       ..........--A---...GC-GCGCAA--AGCAT-GAGAGC--CG-GAGAGA-TC-T--GAGC--G-GG  2556
VER_AGM3      ..........---AT-.....................................................  2550
GRI_AGM677    ..........---C----CC--GAGGCA--A-CAG-GC-TGC-C-A--TC-G--GAGG--T-CGAGG--A  2466
SYK_SYK       CTATGGGTTCA-C--G--C----TCGCA--AGCTTCGC-CGGAAGG-AAA--CGCGATTGG--G--G-CT  2400

ENV_CONSENSUS ?ggac?ga?ag??t?at?gaagt??t?caaagagttgg?agagc?aTtctcaacat?CC?agaaGaaT?a  1893
A_U455        T---TA--T--GG-T--A---A-AGGA----C-A----T-----T-----------A--T--------C-  2542
B_HXB2R       G----A--T--GG-T--A-----AG-A---G---C-T-T-----T----G-C----A--T--------A-  2542
C_UG268A2     G-A--A--T--AA-T--A-----GGGA---G--A----G-----T--C---C----A--T--------A-  2542
D_ELI         G----A--T--AG-T--A---A-AA-A-------C-T-C-----TG----T-----A--C--------A-  2533
E_TN2432      T----A--T-AGG-T--A-----AGCA---G---C-T-G-----C------                      2541
F_BZ163A      G-A--A--T--AA-T--A-C--CTT-G------A----T-----T---T-G-----A--T-C------A-  2491
G_LBV217      T----A--T--GG-C--A-----AGCG-------C-T-C-----T--------T--A--T------G--A-  2566
O_ANT70C      T----T--CG-CA-A--C-C--GGA-A------A-A--A-C--GA----GT-----C--A--G-----T-  2563
O_MVP5180     T----T--CG-CA-C--CTT--GTC-A------A-A--ACA--GAT-C--TC----C--A--------T-  2602
O_VAU         T----T--C--CACA--CTT--GCA-A-----CA-A--G----GG-----T-----A--A-----G--T-  2605
CPZGAB        G-T--A--T--AA-A--A----CTT-T---GTTACACTT---ATT---AGA-----C--T--GC-C--A-  2536
A_ROD         A--...-GCTTG...TGGAGG--AT-GG--C--A-C--G--G-GA--A---GCGG-T--A-----G--C-  2548
B_EHOA        A--...ACCTCG...TGG-G---CC-CAGG----CG-CAG---AG--CA-TGCA--A--C--G--G--C-  2530
SD_MM251      A-A...--CTTA...TGG--GACTC-TAGG----G---A---TGG--C---GCA--C--T--G--G--T-  2617
STM_STM       A-A...--CTTA...TGG---ACTC-TGG---G-----G---AGG--CGGGGC---C--A--GC-C--C-  2620
VER_AGM3      ..............TGGCTT-CTTGCAG-TCC-C-TAT--G--A--CA-----TCT--A------G-GC  2605
GRI_AGM677    GCCTGG-GC--AC-GGGT-CTA-TG-C-G-TCC-C-TATC-G--AG-CA------GT--A------G-GC  2536
SYK_SYK       ATTTGGC-AGCAA-ATAC-CC-CTACCAG-C------TCGAG--CG--GCAGC-T-G--GC-G--GC-TC  2470
```

```
ENV_CONSENSUS GaCAaGGc?t?GAaag??c?tT??t?                    1911
A_U455        --------T-A-----GG-T--GC-A                    2568
B_HXB2R       ----G---T-G-----GATT--GC-A                    2568
C_UG268A2     --------T-T---GCAG-T--GCAA                    2568
D_ELI         ----G---T-A-----GT-T--AC-T                    2559
F_BZ163A      ----G---T-C-----GG-T--GC-A                    2517
G_LBV217      --------T-G-----AG-T--GC-A                    2592
O_ANT70C      ----G---T-A-----AAGT--AT-G                    2589
O_MVP5180     -------TGCA-----AATC--AG-G                    2628
O_VAU         ----G---C-T---C-ACTCC-GT-A                    2631
CPZGAB        -------GC-G-----AG-T--AC-T                    2562
A_ROD         ----G--AGCA----TCG-CC-CC-G                    2574
B_EHOA        -------GGCT---CTCG-CC-CT-G                    2556
SD_MM251      -G-----GC-T--GCTCA-GC-CT-G                    2643
STM_STM       -------GC-T---CTCA-AC-CT-G                    2646
VER_AGM3      --------C-T---G-AATCC-TAAT                    2631
GRI_AGM677    -G------C-T----AAGTCC-GGGG                    2562
SYK_SYK       --------T-G--G-TCTACC-TAAC                    2496
```

```
CONSENSUS-A         ATGAgAGtGAtGggGatacagaggAatTatcaacacTtg...TgGAga??????.....      45
CONSENSUS-B   ATGGCA----------a------cag--a---------g------???-----------??????????      51
CONSENSUS-C         ----------?--------?--------g------a-g-...----T-------?????????.      43
CONSENSUS-D         ----------g--?----g------------g------...-----------??????????      44
CONSENSUS-E         ----------A--a--c-----t-----gg-c-a-----...----a-------.........      45
CONSENSUS-F         ----------G------G----------GG--g------...G---A-.............      45
CONSENSUS-G         ----------a----------------GG---------...----a-.............      45
CONSENSUS-O         ----?-----cTAT--a-G?a-T-g-GA-GA?gA--A?-...AA-Tt-.............      41
CONSENSUS-U         ------------a-------------g--c-----?-...----?-.............      43
CONSENSUS-A         ------------------------------------...-----------.........      45
CONSENSUS-CPZ       ----?????-?-??--???-??????????????????...?T????------???.......  10

                                                        signal peptide \/ gp120
                                                                  <- vpU
CONSENSUS-A   ..??????TGG...gG?actatgaTctTtgGgatgataataATttgT...a?Tgct????cagaa...aA      92
CONSENSUS-B   ??---------...--c--c---c--c--------t-g--g--c---...-g-------g-----...--   101
CONSENSUS-C   ........---...--c-TCT-AGG----t-----t----g------...-a--tg----tg-gg...--      92
CONSENSUS-D   ??---------...--C--c---C--C--------T-g--G--a---...-g--t-...g-----???--      94
CONSENSUS-E   ..---------...--g---t-----C-----t--g-G---------..?-g---c...T----c...--      95
CONSENSUS-F   ........---...--CCT-T-AT--C-G-----AT-g-----C---...-A-----...-??---...--      92
CONSENSUS-G   ........---...--a---T-----C-----T--G-?---------...-g---C...t--a-t...--      94
CONSENSUS-O   ........---AtCTtgT?C?-????a-G-CTT-------CCCA---...TTGAGc...t?TA?c???C-      86
CONSENSUS-U   ........---...----T?t-a----------?----g------...-A---?...at---????--      88
CONSENSUS-A   ..---------...-------------------------------...---------...---...--      92
CONSENSUS-CPZ .....????????-????-???T--G......C---?CC??-?-????-GA??...-????G?????      26

CONSENSUS-A   ?tTGTGGGTcaC?GTcTacTATGGGGTACCTGTgTGGaaaga?gcagag......ACcACccTaTTttGt    153
CONSENSUS-B   a----------a-----t----------------------a---acc.....-----t---------    165
CONSENSUS-C   C----------a-----t----------------------a---A-a??????--t--t---------    156
CONSENSUS-D   t----------A--t--T----------------------A---acc.....--t--t---------    158
CONSENSUS-E   C--------T--A--t--T--------t----------g---t-----t.....---------------    159
CONSENSUS-F   c----------A-----T----------------------a---ACc.....--T--T-----C---    156
CONSENSUS-G   C----------A-----T------------------G----t-----t.....-----t---------    158
CONSENSUS-O   ac-?-AT-C?--A-----T?C------?-----A---G----t---a?a.....C-AgtA?----C---    144
CONSENSUS-U   c-c---------A-----t----------------------t-----?.....-----t---------    151
CONSENSUS-A   ----------------------------------------------.....---------------    153
CONSENSUS-CPZ -???-??--A--A--A-??-----A--?-----?---????-T--????.....C-????--?------    67

CONSENSUS-A   GCATCagATGCtAaAGCatAtgatacaGAaat?CAtAATGTcTGG..?GctACaCAtgCCTGTGTACCcA    219
CONSENSUS-B   ----------------------------gg-a--------t---...--c-------------------    232
CONSENSUS-C   ----------------------g------G-g----------...----------------------    223
CONSENSUS-D   ---------------t-----a-a-----ggca------a-----...--------------------a-    225
CONSENSUS-E   ----------C------C----G------g-G--C---------...--c-------------------    226
CONSENSUS-F   ---------------T-------a-a---gG-A-----------...----------------------    223
CONSENSUS-G   -----T---------------ag---T----gc-----------...----------------------    225
CONSENSUS-O   --T-----------CcT-ACAAGC--T---cAG------A-T---...--At----A---------T--t-    211
CONSENSUS-U   ---------------t-----a-?------?ca-----------...----------------------    216
CONSENSUS-A   ----------------------------------------..----------------------------    219
CONSENSUS-CPZ --C?--?-?--??????????AG------GG?---?---?-?---...--?--???????-----?--?-    106

CONSENSUS-A   CAGACCCcaacCCAcaAGAaaTacattTggaa...AATGTgACAGAAaagTTtaAcATgtGGAaAAATAa    286
CONSENSUS-B   --------------------g--gta------???-------------t-------------------    299
CONSENSUS-C   ----------------------ggt-------...-----a--------t-----------------G-    290
CONSENSUS-D   -----------------------g-ac-----...-----c--------c-------------------    292
CONSENSUS-E   ------------------------cc-----...-----a--------t-------------------    293
CONSENSUS-F   ----T-----T---------G--gt-C-----...-----A--------t---g-T-------------    290
CONSENSUS-G   -------------------g---?c-C--?--...-----A--------t-----?-------------    289
CONSENSUS-O   ----?----cT-----T---TaT-ca?-aa-?...----------T--c--?--T--A---------T-    274
CONSENSUS-U   ---------g-------------?T-------...-----A--------?-------------------    281
CONSENSUS-A   ----------------------------------..----------------------------------    286
CONSENSUS-CPZ ?---T--???T--T???-?-G-????C-????...---?????--??T??-----????-??------?-    141

CONSENSUS-A   cATGGTaGAgCAgATgcAtgaaGATaTaAtcAGtcTATGGGAc...cAaAGcCTAaAgCCATGTgTaaAg    353
CONSENSUS-B   ---------a-----------g------------t-------t...-----------------------a    366
CONSENSUS-C   ------g--t-----------g------------t-------T...-----------------------    357
CONSENSUS-D   ------G--------------g------------t-------t...----------a----------a    359
CONSENSUS-E   -----------------g--g---g--------T-------T..?-----t-----------------    360
CONSENSUS-F   ---------A--A------AC-------------T-------T...--------------------G---    357
CONSENSUS-G   ---------A----------G-----------?t-------T...G---------------------    355
CONSENSUS-H                                                                ------      6
CONSENSUS-O   ---------?--A-----g-----C--t--t---T-?-----?...--G--Tt----a--T-----tc-A    338
CONSENSUS-U   ?--------?----------G------------T-------T...----------a----------?    345
CONSENSUS-A   ----------------------------------------..----------------------------    353
CONSENSUS-CPZ T-----?--????---???-----?--??-?????-?-???-?...--?--?-??-----?--------?    181
```

```
                                  <- tev or tnv in LAI/IIIb
                                 \/ 3'sj in LAI/IIIb
CONSENSUS-A   tTAaCCCCtCTCTGcGTtACttTaaatTGt...a?t??????????????????????????????aa??   387
CONSENSUS-B   --------a-----T-------------c...-c-gat--g-----------------------??--   403
CONSENSUS-C   --g-----A-----T--c--------?---...-ga-----------------.............--tg   393
CONSENSUS-D   --------a-----T--C--------c--c...-c-ga------g---------------------??--   395
CONSENSUS-E   -----T-----------------------...-ccaatgct-----------------------tt   403
CONSENSUS-F   --------A-----T---------?-----???-c-aatgccac----.................--tg   401
CONSENSUS-G   ?-------------T-----------C---...-c-------------.................--tG   391
CONSENSUS-H   C-------------T-----------A---.............................AAA--TT    43
CONSENSUS-O   A-G--TTTC?-g--T--ACAAA-G------...-ca-----------------.........------A   373
CONSENSUS-U   -----------?--T--?--------C---...-c-gatg-aaag---------.........-----Ta   388
CONSENSUS-A   ----------------------------...-------------------------------------   387
CONSENSUS-CPZ ?-?--?---?-?--T?-A-???-G?-?--?...--?-----------------------.........   200

CONSENSUS-A   tcac??????????????????????????????aa?a?cac?a?????????????????aa?a?   401
CONSENSUS-B   ????-----------------------------tg-t--t?-t-------------------t--   416
CONSENSUS-C   ?t-----t-------------------...-----tg-t--c?-ta-----------------t?-   410
CONSENSUS-D   ??--------------------.....--tg----t?--------------------t?-   408
CONSENSUS-E   -g--c---------------------aatg-ca-t--c-ta--c-atgtc-ctaaca-aatagga--t-t   448
CONSENSUS-F   ?---------------------a--g-cact---??-?-???-?--caa-ac-g-.....??-?-   416
CONSENSUS-G   -a--ctatact------.................??-?-???-.....................-A   404
CONSENSUS-H   AT--TAGACCT.....................--TTGT--TGGAACT.............--T-G    74
CONSENSUS-O   -a--a-----------------.....--tga?--a--a---.............??-?-   388
CONSENSUS-U   ga--tAac--------------.....--T?-?--tgaAg-a.............??-?-   407
CONSENSUS-A   ----------------------------------------------------------------   401
CONSENSUS-CPZ .....................................-A???A-CA---------------.....   206

CONSENSUS-A   ?ac??????????.....????????????????????ga?.....................   405
CONSENSUS-B   --?t---------?????????-----------a--------a?-??????.....?????????   420
CONSENSUS-C   t--c---------???.....-----------------??-.....................   414
CONSENSUS-D   -------------???.....---------------??-????????????????????????   410
CONSENSUS-E   a--a------.....................................????????   452
CONSENSUS-F   -??----------.....................................???accctg   422
CONSENSUS-G   TtgtACT------???g?gaac------aacAAtagcactgTg??-.....................g??   432
CONSENSUS-H   T--TGTAAGTAATAGTACTGAAACTAATTGTACTGGAAAGAACA-TAGCTTTGTGGATGATCGTATC...   141
CONSENSUS-O   -??----------?????????--------------------??-??????.....   388
CONSENSUS-U   -??-------.....------------------.....   407
CONSENSUS-A   -------------.....----------.....   405
CONSENSUS-CPZ .....---A-----A---C----?C-CC?????????????.....   212

CONSENSUS-A   ...atgaaa???...g?agaa.............ataAaaAAcTGctCTTtCAATAtgACcaCAgaa?   449
CONSENSUS-B   ???---g-gaaa???-g----?????????????----------c------agca   469
CONSENSUS-C   ?????????---???------.....?????????--------t----------gca----------   452
CONSENSUS-D   ??????g-?---...a---g-.............--G-------------a---------g   451
CONSENSUS-E   ?????????.....-at---.............G---g------T-----T----------c   492
CONSENSUS-F   aaggaag-?cca...-gg-c-.............---C-------T-------------------g   473
CONSENSUS-G   ??c-aa???.....-a----.............--g-------------------A---------a   476
CONSENSUS-H   .............CA----.............T----------------------A---------A   181
CONSENSUS-O   .............aat--g?????.....?-G---?-g--Tgag--T---g-A--?----Ttc   425
CONSENSUS-U   .....???---???-----?.............--g-----?----------?-A----------   442
CONSENSUS-A   ...---------...------.....----------   449
CONSENSUS-CPZ ?????????---????T??-?.....?-??--?-?--T???--T---??--??----?-T   235

                 -> tev or tnv in LAI/IIIb
              5'sj \/
CONSENSUS-A   TAAgagATAagaaacagaaagtatattCaCTtTTtTAtAgacttGATgTaGTaCaAAttaat????????   511
CONSENSUS-B   ----------gtg-------a----g----------a----------c---ag--aat-at--   536
CONSENSUS-C   ----------------------g---g-----------------a------c-c----------   514
CONSENSUS-D   -------------ga--c-----c--g----------a---------?--------ag--?--------   511
CONSENSUS-E   -------------g-----G--cc--G----------ag------a----------g-agataa---   559
CONSENSUS-F   ----------C-G-t-------C--g--------------------A------c---c-gcaat---a-   539
CONSENSUS-G   -------------a---?--A---CG-g-----C----a----------C---------------   537
CONSENSUS-H   -------------GA---C--A---CG-T-----C-------A----------C-------GGAAAG..   249
CONSENSUS-O   -C-a---c--A--GG-----AA-C-GG-T--A--C---GT-tCA---t-g?-ta-?------gacacaa-   492
CONSENSUS-U   -------------Ga---?--a-?--g-?-----?----a---a----------c-------gataat--   506
CONSENSUS-A   ----------------------------------------------------------------   511
CONSENSUS-CPZ -?-?---------??-A?-??-??-??-??-A------???G??---?-??-?A-??????G---.....   269
```



```
CONSENSUS-A   tacaAga...Aaaagt????????gtacgtata????????GgacCAGGaCaA?????gcatTctaT   900
CONSENSUS-B   -------...-----------...a---a-----------------gag-.....-----t---   919
CONSENSUS-C   -------...------.........a--a-g---.........----------.....a-------   906
CONSENSUS-D   -------...c----?.....---ac--c----.........----t---g---.....---c-----   897
CONSENSUS-E   -------...-C----.........a--ac----.........----------.....-t-------   946
CONSENSUS-F   -------...------.........A---a----.........----------g-.....-----t---   931
CONSENSUS-G   -------...------.........A--a??t-c.........----------.....--g------   931
CONSENSUS-H   -------...-----------...A--a-g--c.........----------.....--g------   635
CONSENSUS-O   ?---gT-...C--gAg.........A--a??---.........--T---...aTG.....--?-GG--c   859
CONSENSUS-U   -------???------.........a---------.........----------.....---------   880
CONSENSUS-A   -------...-------------------------------------------------------   900
CONSENSUS-CPZ ?---???...?G-?A?.........?-?-A?---.........--?------ATG......A-?--?---   499

CONSENSUS-A   gcA...acaggtgacata..............atAGGg...gAtAtAAgacAaGCacAtTGtaatgTcA   949
CONSENSUS-B   a--...-----a?-a---????????.....-----a...----------------------ca-t-   967
CONSENSUS-C   ---...-----a------.............-----a...--c--------------------cA-T-   955
CONSENSUS-D   A--???---???aga---????????.....-----a???----------------------a-T-   943
CONSENSUS-E   ag-...-----A------.............-----a...---------a-----T-----G-gA-T-   995
CONSENSUS-F   ---...-----a--a---.............-----a...--c-c---a-G----------c--t-   980
CONSENSUS-G   ---...-----------..............-----a...-------------------------t-   980
CONSENSUS-H   ---...--------?---.............-----A...-------------------------T-   683
CONSENSUS-O   AGc...-TG--c...c--gag?ga???aa?aca-a-??????agctC---Gg?---TT-------?TAT-   910
CONSENSUS-U   ---...--------a---.............-----a...----------------------?a-t-   928
CONSENSUS-A   ---...-----------..............------...-------------------------   949
CONSENSUS-CPZ AA?...?T--AAA??G?-.............??---A...--?-C?--????--?T?C----???-?-   529

CONSENSUS-A   gtaga?cagaaTGGaAtaaaacttTacaa?aggta...........gcta?acAatTAagaaaa.....   999
CONSENSUS-B   -----g--a---------c------a--c--a--...........-t----A-------g--?????  1019
CONSENSUS-C   ---a-ga?a--------?----------a-----...........ag--a-a-----gc-g--.....  1006
CONSENSUS-D   --g?-g----------------------c-----...........----a-A-----g--g-c?....   995
CONSENSUS-E   A-g--A--a--------g--g-----a--c-----...........a--ga-a------a-g-g.....  1048
CONSENSUS-F   --g--a--C---------?--?---g--a-----...........a?ggc-a-g----agtct.....  1030
CONSENSUS-G   -----a?-a-t-----?g-g-tG-----ga-t--?...........a??gc-?--C---a-g--.....  1027
CONSENSUS-H   -----g?-a-?------g-g-tg---?-?a-----...........-?-?c-------?a----.....   729
CONSENSUS-O   a-gccA-T--t---g-a---g-c---a--c-aAc-...........---ga-AGg-ATTT-g--.....   963
CONSENSUS-U   --g--a----------?---------c-----...........------?----------.....     978
CONSENSUS-A   ----------------------------------...........----------------.....     999
CONSENSUS-CPZ A-??GA-???----?-?C?-?-?????-?--A??-??????????????AG??T-????--??.....   552

CONSENSUS-A   .tacTtt?????????aacaaaaca...........??????ataatcTTtgctaac...?cctcaGGa  1037
CONSENSUS-B   ?c-a---...g-g-----t------...........---------g-----aa-c-a???t--------  1060
CONSENSUS-C   .c----ccct--------T------...................----aa-----acca...t--------  1048
CONSENSUS-D   .cTtc--......----------.............aca-----t---aaacCa...t--------  1037
CONSENSUS-E   .C---------...a-t--T--G---.................---------caaCCA???c--------  1089
CONSENSUS-F   .c-t--c---...-----tgc----..............-------aa---aactcA...t--------  1069
CONSENSUS-G   .at-------------------?c.........???----------c----aaCtCA...t-tg-----  1065
CONSENSUS-H   .?--------...a----t----??.....-------c----aaacca???t--------   766
CONSENSUS-O   .CTtg-A.........---?-t---??????aata?tgTtAcc--g-?A--Caa-c-?agcAgt???--t  1010
CONSENSUS-U   .c-----.....aat--t-??---..............--------?---aa??cA???t--------  1015
CONSENSUS-A   .-------------------------.....-------------------...---------     1037
CONSENSUS-CPZ .??A???.........--???---?.........???GC-AA--?--?A???A?????...-?-???-??   569

CONSENSUS-A   GGGGAt?TaGAAaTtacAAcacAtAgttTTAaTTGTggAgGagaattt..?TTcTAtTGcaataCAtCag  1103
CONSENSUS-B   -----cCc-------gt--tg--c----------------g------...-----c--t------a--c  1127
CONSENSUS-C   -----cc-----------------c---------a----------...------------------a  1115
CONSENSUS-D   -----ccc---------------c----------------g------...-----C------------a  1104
CONSENSUS-E   --a---C------------tg---ca----------A----g------...---------------a--c  1156
CONSENSUS-F   -----CC------------tg---------------a----------...-----C------------  1136
CONSENSUS-G   -----cC-------------------------a----------...--------t----------  1132
CONSENSUS-H   -----Cc-----------------?----------a----------...--------t------a--a   832
CONSENSUS-O   --A---gc---gg-A--ccATtT?ca------C---CAt--------C...--T-----t--C-----T?  1075
CONSENSUS-U   ------------------------c----------------------...----------------?  1080
CONSENSUS-A   ------------------------------------------------..-----------------  1103
CONSENSUS-CPZ --A---CC?--?G-?-?-???---??G--------???---------?...--?-----T?---??A??C   617

CONSENSUS-A   g?cTgTTt...aatagcacttgg?a??????????????????.....aatagcact?????  1132
CONSENSUS-B   aa------...-----t-------?-----------------????????????---g?t--------  1155
CONSENSUS-C   ac------...-----t--a-ac-?-----------------...?????????---g-t--a-a---  1145
CONSENSUS-D   -a------...-----t--a----?-----------------????????????----?t--?-----  1131
CONSENSUS-E   aa------...----at-----catagga------.....??---gaa--catg--  1194
CONSENSUS-F   aa------...---ga---a???-?----------.....---?a---------  1161
CONSENSUS-G   -a------???----at-g-?t?-?-------.....----at---ac---  1161
CONSENSUS-H   aa------...-----T-------t---t------.....--?--?----a---   862
CONSENSUS-O   -ga-----...---tAT--c-TTtcatgt----------????????????---?-t--ca--t-  1111
CONSENSUS-U   aat-----...-----t-?-----?----.....--????--?-ggga  1107
CONSENSUS-A   --------...-----------------------.....----------  1132
CONSENSUS-CPZ ?-???--?...-?-G?-??????-?-------.....??C-T?--A.....   629
```

```
CONSENSUS-A   ??????????????ca?????????aat?acac????????????.....???????a?tcaaatgac  1151
CONSENSUS-B   --------------??-----------a-?--t-----------??????-------?-????-?--?  1166
CONSENSUS-C   --------------??----------?-??----------??????-------?-?????????    1149
CONSENSUS-D   -------...----??---.....---a?t--a-----------??????-------?-????-????  1140
CONSENSUS-E   ----------gaggggtgt...-----gg---t.....................?????????    1212
CONSENSUS-F   -----.................???-????----------.....-------?-?????????    1161
CONSENSUS-G   --------------??---...-------??-----------.....----?-?????????    1166
CONSENSUS-H   ----...aat----??---.....---g??---.....................-?-??????a??   872
CONSENSUS-O   --------------??-aa---a-c---gt?--taAt--t.....................a??----g-  1134
CONSENSUS-U   -------.....-??---...-------?--?-------...---??????-------?-???-?--?t  1116
CONSENSUS-A   ------------------------------------------.....----------------------  1151
CONSENSUS-CPZ ..........................................................--?-G?     633

CONSENSUS-A   actATaact......ctccaa...TGcAgaATA...AagcaaaTTgTaaatATg...TGgCAgagagtaG  1206
CONSENSUS-B   -----c--a??????----c-...---------...--a------a----c---...------ga-----  1221
CONSENSUS-C   -?c--c--a......?---c-...---------...--a------a----c---...------gag----  1202
CONSENSUS-D   --a--c--a.....----c-...---------...--A------a----c---..?------g------  1195
CONSENSUS-E   ???--C--A......--T-C-...----ag---..?---------a----c---...------g---c--  1264
CONSENSUS-F   -----c---.....----c-..?--T------...--a----------c---...------ga---g-  1216
CONSENSUS-G   ???--c--a.....----c-...--T--g---...--A---------ga---...----------g-  1218
CONSENSUS-H   -??--t--a.....--g-c-...---------...--A--------g-ga---...----------g-  925
CONSENSUS-O   ---?a????.....a-A-CT...-----gt---...-gA---g-g----Gatca...---AT---G-G--  1184
CONSENSUS-U   -?---?--a.....----C-...---------...--a---------g?---...-----------  1168
CONSENSUS-A   ---------.....------...---------...----------...----------...-------  1206
CONSENSUS-CPZ ????-?-TA......???-??...--------T...-??--G--A--T-?-???...---???-??--??  665

CONSENSUS-A   gacaA...gCaatgTAtgCccCtCCcATccaa...gGagtaaTaaggTgTgaatCAAacATtACaGGacT  1270
CONSENSUS-B   --a--...---------------------ag-...---ca---t--a---tc------t--------g--  1285
CONSENSUS-C   ---g-???---------------------tg--...---aac----ca---a-------t--c--------  1266
CONSENSUS-D   --a--...---------------------g--...---c----c-aa---tc------t--------g--  1259
CONSENSUS-E   -----...----------T---------agt...---ag---T-At----T------T---------A-  1328
CONSENSUS-F   ---g-...------------g-------tgc-...---aac--T-cc---A?c-----t--------t--  1279
CONSENSUS-G   -----...---------------------GC-...---AaC--t-ca---a------------------  1282
CONSENSUS-H   -----...---------------------gc-???---AAc--T-cA---a------------------  989
CONSENSUS-O   -gTCG...-G-c-c-----a----------?-...--tAAcc---cA---a??--------A--T---A-  1245
CONSENSUS-U   -----...---------------------tgc-...---a?---t-??--c?c------t----------  1228
CONSENSUS-A   -----...---------------------...------------------------------------  1270
CONSENSUS-CPZ ?-A?-...-G?--?--???A?-C--?-??AG-...--?AA??-??C?--?A?T--?-?---?--T---?-  706

CONSENSUS-A   acTatTaACaagaGAtggtGgg?at???.............aataata?ta?a???.........??????  1304
CONSENSUS-B   g----------------------ta-----????????????--?--?-cc?-?---???...???------  1317
CONSENSUS-C   ------g---c-t-----a--aa?a---????????????-----?-ca?-?---????????------  1298
CONSENSUS-D   -t----g--------------ta?a---???.........----?----?-?---......???------  1290
CONSENSUS-E   ------g--------------tgc----???.........-------cg-c?---.............  1364
CONSENSUS-F   ------g--------------tc-gaat??????.....---?-?-c-?-?---???.............  1314
CONSENSUS-G   ----------------------a-----.....-------ca?-?.....................  1316
CONSENSUS-H   -a---------------------??-a-???.....-----??c-?c?---.................  1020
CONSENSUS-O   GA-tc--CA-?Tg---a??...cCATGG.............--c-?c-gc?-?---.............  1275
CONSENSUS-U   -?----g--------------ta-----?.............----?--g-?-?---.............  1260
CONSENSUS-A   ------------------------.....-----------------------.....-----------  1304
CONSENSUS-CPZ T?-?--???????--???-C??-T?--T.............--??-?---.....................  721

CONSENSUS-A   ?????????aatgagAccTTcAGaCCtggaGGagGAgatAtgAgggAcAAtTGGAgAAgTGAatTATAtA  1365
CONSENSUS-B   ----------c?----t-------------------------------------------------------  1377
CONSENSUS-C   ---------???-----a-------------------------------------------------------  1356
CONSENSUS-D   -------------------------------------------------------------------------  1351
CONSENSUS-E   ......-----c----------------------A----a-A-------------------------------  1425
CONSENSUS-F   ...------???----------------g------a------aa-----------------------------  1372
CONSENSUS-G   ...a----------------------------------------------------------------------  1378
CONSENSUS-H   ......---???-------------------------------------------------------------  1078
CONSENSUS-O   -------a---c???--ac-T-----a-t---G----------aA--T-TA-------C------g-tC-  1334
CONSENSUS-U   ...----a--?---------------------------------a-----------------------------  1321
CONSENSUS-A   ----------------------------------------------------------------------  1365
CONSENSUS-CPZ ......---????T?-?A?-???----???-??-?-A???-???A--????------???--G?-?-??-  749

CONSENSUS-A   agTAtAAaGtaGTaaaaaTtgaaCCacTaGGagTAgcACCcacCcgg...GCAAaaaGaAGAGtGGTG..  1430
CONSENSUS-B   -a----------------------t----------------aa-...-----g----------..  1442
CONSENSUS-C   -a--------g---g-----a-g---t-g--------------ta?-...--------g---------??  1420
CONSENSUS-D   -A----------g-------------------------a--...-----g----------..  1416
CONSENSUS-E   -A----------C----------------a----------a--...-----G----------..  1490
CONSENSUS-F   -a----------g-------------------------A--...---------ca-------..  1437
CONSENSUS-G   -------------------A----------------------a--...----GG----------..  1443
CONSENSUS-H   -?----------------c?--?----------------a--...----gg----------..  1140
CONSENSUS-O   -a--c----------ggg-AA----TT-TA-T--g-----t--AaaaATt----GgCC--tcA-A-gcAc  1404
CONSENSUS-U   -a----------------------?----------a--...-----g----------..  1385
CONSENSUS-A   ------------------------------------...----------..  1430
CONSENSUS-CPZ ?---??-?--?--????--?-??--?T-?TC---?--?--A--AA?AG??A?--GG????C--????AA?  790
```

```
                                gp120 \/ gp41                                  <- RRE or CAR
CONSENSUS-A   .........gagAGAGAaAAAAGAGCa??????gTtGgA...cTgGGAGctgTctTCcTTgGGTtcTTa   1481
CONSENSUS-B   .........c-------------------.....--g---???a-a------a-g-------------g   1493
CONSENSUS-C   ????.....-------------------...-----g---???a-a--------g-------------g   1471
CONSENSUS-D   ..........--A----------------.....A-A---...--a------a-G-------------G   1467
CONSENSUS-E   .........--------------------.....--G---...A-A------A-Ga--t-----------   1541
CONSENSUS-F   .........a-----------------------g---...a--------t-g-------------g   1488
CONSENSUS-G   .........-G-------------------.....------...?-------------------------   1493
CONSENSUS-H   .........--------------------.....------...-------------------------   1191
CONSENSUS-O   T?g?aCt???C-t----------------.....--A---...T-----ATGC-A---T-G---g-tC--   1460
CONSENSUS-U   .........--------------------...-----g---...a--------?-G-------------G   1435
CONSENSUS-A   .........-------------------------------...-------------------------   1481
CONSENSUS-CPZ ACA??A?...??C---C-?-------??...---?-?--?...?-?--?????-?---?-?--??-?C-?   823

                                  /<-      RRE primary binding site      ->/
CONSENSUS-A   GGAgCaGCaGGAagcactatgggcgcggcgtcaataacgctgacggtacaggccagacaattattgtctg   1551
CONSENSUS-B   -------------------------a--------g-------------------------------------   1563
CONSENSUS-C   -------------------------?-------------------------------------?-------   1539
CONSENSUS-D   -----------------g--------A------?-g-----------------------g----------   1536
CONSENSUS-E   ---------------------------------------------------------------------   1611
CONSENSUS-F   ---------------------------------------------------------------------   1558
CONSENSUS-G   -------------------------a-----------------------T-----------------   1563
CONSENSUS-H   ---------------------------------------------------?-----------------   1260
CONSENSUS-O   A-T--------t--------------A---G---C-g--------------a--CAcacT--ga--AAG-   1530
CONSENSUS-U   -------------------------?---------------------------------------------   1504
CONSENSUS-A   ---------------------------------------------------------------------   1551
CONSENSUS-CPZ ?-T-----?--?--?--?--------A---?--?--?---------?----?--?-??-?--G?-????-   874

CONSENSUS-A   gtatagtgcaacagcagagcaatttgctgagggctatagaggctcaacaacatctgttgaaactcacggt   1621
CONSENSUS-B   ------------------a-----------------t-----g-----g---------c-------a--   1633
CONSENSUS-C   ----------------a--------------?-----------g-----g---a-----c----------   1608
CONSENSUS-D   ----------------A-A-----------------------G-----g---------C-------?--   1605
CONSENSUS-E   ----------------a-------------------------g--g--g---------c-------a--   1681
CONSENSUS-F   -?----------------a-----------------t--a--G-----G---------C-g-----a--   1627
CONSENSUS-G   -C--------------A--------------a-----------G--g--g---------C-------A--   1633
CONSENSUS-H   -c----------------a-----------------------?--?--g---------?-------a?.   1325
CONSENSUS-O   ----------------GA---CC----a--A--a---C----C--G-----aT--C---GG--AT-T--   1600
CONSENSUS-U   ----------?-----A?A?----------------------G-----G---?-----C-------A--   1570
CONSENSUS-A   ---------------------------------------------------------------------   1621
CONSENSUS-CPZ ----?--?--?--?---???---C--?--?AA--?------?-?--?--?--?--??-?C-?--?T-??-   922

                                                          RRE or CAR ->/
CONSENSUS-A   ctggggcattaaacagctccaggcaagagtcctggct?tggaaagatacctaa?ggat...caacagctc   1686
CONSENSUS-B   ---------c--g-----------------------g--------------a----..?---------   1700
CONSENSUS-C   ------------g---------a--------------a-a------------a----...---------   1675
CONSENSUS-D   ----------------------------?--------g--------------a----...---------   1671
CONSENSUS-E   ---------------------------------------g--------------a----...---a--t--   1748
CONSENSUS-F   ---------------------------------------g--------------a----...---------   1694
CONSENSUS-G   ---------------------------------------g-A------------a----...---------   1700
CONSENSUS-H   .....................................................................   1325
CONSENSUS-O   A-----T--?-G---A----GA--TC-CC-G--A--CT-A----CC-T?a--CA-A--...--G--A---   1665
CONSENSUS-U   ----------------------------?--------G-------------?A-??-...---------   1633
CONSENSUS-A   ---------------------------------------------------...---------   1686
CONSENSUS-CPZ ------AG-A-----A--?--?--?---?-G--T--?G-?--?-?G-----??-?---...--?--??-?   973

CONSENSUS-A   ctaggaatttggggctgctctggaaaactcatctgcaccacta?tgt?ccctggaactctagttgg....   1750
CONSENSUS-B   --?--g--------t-----------------t---------gc---g--t-----tg--------....   1765
CONSENSUS-C   -----g-------------------------------------gc---g--t--------?------....   1740
CONSENSUS-D   --------------t----------?--a---t--------------g-----------------....   1735
CONSENSUS-E   ------c---------------------a--------------gc---g-----------c-c----....   1814
CONSENSUS-F   -----gc-----------------------------------a---g-----------------....   1760
CONSENSUS-G   -----g-----------------------------------A---G---------a--------....   1766
CONSENSUS-H   .....................................................................   1325
CONSENSUS-O   ---AaCC-?--------tAA?----??--A------TA---aTCA--AaAa-----?AaA-CA---acAg   1730
CONSENSUS-U   -----??-------?---?---------?---?----?---------G--------------C---....   1691
CONSENSUS-A   ---------------------------------------------------------------....   1750
CONSENSUS-CPZ ?-G?-CC-?-----?--??---??--GG????---T?A?--?-CG--G--T-----?AA?TC?---????   1021
```

```
CONSENSUS-A   .....................agt...aataaa?ct?a?a?tga?atatggga?aacatgacctggct  1788
CONSENSUS-B   .....................---???------t--ctgga-?-g--t---?-t-----------a-  1807
CONSENSUS-C   .....................---???----?-t--c-a?-?--t--t---?-t-----------a-  1779
CONSENSUS-D   .....................---...----g-t--cta?a---g--t---c-------------a-  1777
CONSENSUS-E   .....................---...----g-t--tttgaa--g--t---a-c--------a---a-  1858
CONSENSUS-F   .....................---...------t--c-ggag--g--t----gg-----------a-  1804
CONSENSUS-G   .....................---...------t--TtT-A---g--t-----t-----------A-  1810
CONSENSUS-H   ..................................................................... 1325
CONSENSUS-O   ga...................????????????ga-a-tGAaagt--T-----c---?-a--A----A  1766
CONSENSUS-U   .....................---...------T-?-T-G-???-?-T---?----??--------A-  1722
CONSENSUS-A   .....................---...---------------------------------------- 1788
CONSENSUS-CPZ ?????????????????????-?C...-?????-??A?A?A???---T----?---T?-?--A----A  1046

CONSENSUS-A   gcaatgggataaagaaattagcaattacaca?a?ataatatat??t...ctaattgaagaatcgcagaac  1851
CONSENSUS-B   -g-g-----a-g-------ga----------agct-------cacc...t------------------  1874
CONSENSUS-C   ---g-------g---------t-----?-------c------cagg..?t-gc-------c-----a---  1843
CONSENSUS-D   -g-g-----a-g-------ga----------gg-t-------?agc...t----------------a-?-  1841
CONSENSUS-E   ag-------g-g-----------------a-cca-------gag...a--c--ac-------------  1925
CONSENSUS-F   -g-g-----a-----g------------t--a-cga------cagg...t----------?---------  1870
CONSENSUS-G   Ag-------?-Gg-------A----------c-cCa------CaaC...--Gc---------------?-  1875
CONSENSUS-H   ..................................................................... 1325
CONSENSUS-O   -?--------C-?C-g--A-a---c?TA-GctcC-?c------GA-...gA---AC--?-gG-A--AG?a  1827
CONSENSUS-U   -G-G-----?-G-------?----?-------?-G-?-----?--?...T-??-------?--?--A-??  1772
CONSENSUS-A   ----------------------------------------...---------------------- 1851
CONSENSUS-CPZ ??-------?-?-TT-G-????--?-??---GG-?A?--?-?----...?-?T-A?-???-G-A--????  1085

CONSENSUS-A   cagcaggaaaagaatgaacaagacttattggcattggacaagtgggcaa?tctgtggaattggtttgaca  1920
CONSENSUS-B   --a--a----------------a-------a---?--t--?-------g-t--------------a---  1942
CONSENSUS-C   --------?--a-------a----t---?-a----------??---?a--a---------?--------??--  1905
CONSENSUS-D   -----a-----------------a?-----?a----------a-------g-t--------------a?--  1908
CONSENSUS-E   --------c-g-------a-g--t--g--a-a------t--a-------g--------------------  1995
CONSENSUS-F   ----------------------a-----------------a-------g--------------------  1940
CONSENSUS-G   ------------------------------------------G--------g-------Ag--  1945
CONSENSUS-H   ..................................................................... 1325
CONSENSUS-O   --?------?-a-----ga--?-?--Gc---AG--A--TG-A-----CTC-a-T----------c------  1893
CONSENSUS-U   -----?-------?----?----?---?-??A------?--?-------G-?-------?-------?--  1830
CONSENSUS-A   ----------------------------------------------------------------- 1920
CONSENSUS-CPZ --A--A--G-?A---?--A?G--??---??-A-?-?--??-A---??C?-?T-?---------------  1138

CONSENSUS-A   ta?caaa?tggct?tggtatataa?aatatttataatgatagtaggagg?ttaatagg?ttaagaatagt  1984
CONSENSUS-B   --a----------g---------a------c----------------c--gg----t----------  2011
CONSENSUS-C   --a----------g---------a------c----------------c--g-----t---------a-  1974
CONSENSUS-D   --a----a-----g---------a------c----------------c--g-----t----------  1978
CONSENSUS-E   --a----t-----g---------a----------------------t--------t---------a-  2065
CONSENSUS-F   --t----------g---------a------c----------------c--g-----c----------  2009
CONSENSUS-G   --t----c-----A---------g----------------------t--------T----a------  2015
CONSENSUS-H   ..................................................................... 1325
CONSENSUS-O   --A-T--A---T-G----------A----GC?-----C----------CAC--?----t?-----G-t?-  1959
CONSENSUS-U   ------?----?-G----------A------C----------------C--G---------------  1895
CONSENSUS-A   ----------------------------------------------------------------- 1984
CONSENSUS-CPZ --A--?-A---------------A---?--??-T---??--------?-A-??-----?-------T?-  1194

                                3'sj (minor) \/  \/ 3'sj (minor)            \/ 3'sj (major)
CONSENSUS-A   ttttgctgtgctt?ct?taataaatagagttaggcagggatactcacct?tgtc?tt?cagaccc?tacc  2048
CONSENSUS-B   ---------a---t--a--g-g-------------------------at-a--g--t--------cct-  2080
CONSENSUS-C   ------------?t--a--g-g-----------------------?t-?--g--t-------t--?-  2040
CONSENSUS-D   -------------t--t--g----------------------t------c----a--t-------tcct-  2048
CONSENSUS-E   -------------t--a--g-g-------------------------t----t--c--------c-?--  2134
CONSENSUS-F   -------------t--a--g-------------a--------------t----a--a-------a--t-  2079
CONSENSUS-G   -------------T--A--G-----------------------------t-a--a--C-------T-?--  2084
CONSENSUS-H   ..................................................................... 1325
CONSENSUS-O   ?A-GATA?-A---AA-c--G-g--?-ACA------------TCA---CC-C--G-------t--Cc-t-  2025
CONSENSUS-U   ----?-----?--T-----?-?----?--------------------?T---------------T???-  1953
CONSENSUS-A   ----------------------------------------------------------------- 2048
CONSENSUS-CPZ ??-???A?-??-?-??-???-?-G?-??----------?--????---C-?----------?-----?-  1233

CONSENSUS-A   c???acc??aggg?????c?cgacaggcccg?aagaatcgaagaag?aggtggagagcaagacagagaca  2105
CONSENSUS-B   -cagc--cg----ga...-c----------a-g----------a----------ag----------  2147
CONSENSUS-C   -caa---cg----ga----c---------t--g-?----------a----------?-----  2105
CONSENSUS-D   -cagc--cg----ga...-c----------a-g----a-------a----------?--g-----?--  2113
CONSENSUS-E   -atc-t-ag----aa...-c------a----a----------g------c-------g--------  2201
CONSENSUS-F   -caag--cg----aa...-c----------a-g----------g----------g--a-----  2146
CONSENSUS-G   -ACC---AG----aA...-C---------?-A-g----------G----------  2150
CONSENSUS-H   ..................................................................... 1325
CONSENSUS-O   -aaC---a-g?--AA...GcA-?A-C---A-G-----CA-G-----g---------aG-------GCc--  2089
CONSENSUS-U   -CAA?--CG----GA...-C?--------?-G-?---?--------A---------??------???---  2010
CONSENSUS-A   ----------------------------------------------------------------- 2105
CONSENSUS-CPZ --A-?--AG??--A-...--A-????--??---GA---????----------?????-------?A?--  1274
```

```
CONSENSUS-A   gatc?at?cg?ttagtgagcggattcttagcacttgcctgggacga?ctgcggagcctgtgcctcttcag  2171
CONSENSUS-B   ----cggt--a-------at--------------at---------t----------------------  2217
CONSENSUS-C   ----c--t--a-------?-------------------------------------------?-----  2172
CONSENSUS-D   ----c--t--a--g----a--------c-------at-------------a---a--------------  2182
CONSENSUS-E   ----cg-g--a-------------------------------t--a--------------------  2271
CONSENSUS-F   ----cg-ga-a-------a------------t----t---------c---a---a--------------  2216
CONSENSUS-G   ----C--t--a-------------------?-------------c----------------------  2219
CONSENSUS-H   ..........................................................................  1325
CONSENSUS-O   -g-gG-?a-Cc-?GCC-CaA--?-----GC?---gtTG-ACACG--c--CA---CAA-AAT-T-G-GG-?  2154
CONSENSUS-U   ----C??T--A----?--?-----------------?------?--C---------------?-------  2073
CONSENSUS-A   ----------------------------------------------------------------------  2171
CONSENSUS-CPZ -?-?-??G??---???-?A?--G-?---??----??T?--------C--C--A-A----G-GA-T-GG--  1325

CONSENSUS-A   ctaccaccg?ttgagagactt??tcttgatt...g?a...gcgagg...?ctgtggaacttctgggac?c  2226
CONSENSUS-B   ---------c----------ac--------?..-t-?..------?..at----------------g-  2278
CONSENSUS-C   ---------a-------------a------...-c-...-----a...g-?----------------?  2228
CONSENSUS-D   ---------c----------aa----a---...-c-...------...at----------------g-  2243
CONSENSUS-E   ---------c----------ca--------...-c-...------...a-----------------a-  2332
CONSENSUS-F   -----g--ac----------ca-a--a---...-c-...------...at----???...........  2262
CONSENSUS-G   --------------------CA--------...-c-...------...A-a---------------G-  2279
CONSENSUS-H   ..........................................................................  1325
CONSENSUS-O   T-------TC-----CA----Agca-CAGgG...ATCCaGAa-gt-ATCagc?AcCTGgGa--t----Tg  2220
CONSENSUS-U   -----?---------------A--?------...---...??????...-??-???-????---?--?-?  2109
CONSENSUS-A   ----------------------------...---...------...----------------------  2226
CONSENSUS-CPZ ------GA-C----C-AG---AGC--GC-A-...-TG...TG---A...CAAC-CA-GACGT------AT  1386

CONSENSUS-A   a??????????????t??g???ggggtggga?g??ctcaa?tatctg...tGgAAtCTtcTg?taTatT  2268
CONSENSUS-B   -gg...---------...-?---?--------a-cc-----a---tg-...--------c---cag----  2324
CONSENSUS-C   -gcag-ctcagggga--acagag---------a?-c--t--g------...G-A-g----G--Cag----  2291
CONSENSUS-D   ...................ag---------a-cc-----a------...--------C---cAg----  2289
CONSENSUS-E   -gcagtctcaagggac-ga-acg---------a-gc-----a-------...G----------t------  2399
CONSENSUS-F   .......-acagggggc-ga-gag---------a-cc-----act----...G-------CaC-C-G----  2322
CONSENSUS-G   -aCAttCTCaAGGGAC-GA-ACT---------G-GC-----A------...----------t-G----  2346
CONSENSUS-H   ..................................................  1325
CONSENSUS-O   TGGATCcTaGGg.........CA-Aa-A?AATT?a-GcTTGcAGA?-tTgTgcAGC-g-AA?aCA---c-  2276
CONSENSUS-U   ?---------------?--?--G---------A?C---?--A--?---...G-?-?----G--C?G----  2147
CONSENSUS-A   ----------------------------------------------...-------------------  2268
CONSENSUS-CPZ CTAATACTTCACAGCC-CA-ACT-CTAC----GAGA--GTGCCTG---...G--GGAA--A-TCA-----  1453

                              <- rev
CONSENSUS-A   GGGgtCgGGAaCTaAAaA?TAGTGCTattAaTTTGcTTgAtaCcatAGCAaTAGcAgTAGCTggctGGAc  2337
CONSENSUS-B   --a---A---------g-a-------g---gc-----ca--g---c---t----------Agg----  2394
CONSENSUS-C   ------t---g-------aG----------g-c-------------------------AaG-A--  2361
CONSENSUS-D   --a?--a------g--g-a----------Gc-------------------------Agg----  2358
CONSENSUS-E   ----c-A----------T----------TC-------a--g-t-C----------G--G-----  2469
CONSENSUS-F   ------A---------G-A----------GC------A------C---------t--------AGg----  2392
CONSENSUS-G   ------a---g-----G-A----------?-------A----------AA------  2415
CONSENSUS-O   --cTA-AA---T-Gc---a--------CA-??c-?--a-----tc-T---G-G-----t--cAAT-----  2343
CONSENSUS-U   ------?---------?-A?----------G-?--?--?--G---C--------?--------A?G-A--  2209
CONSENSUS-A   ----------------------------------------------------------------------  2337
CONSENSUS-CPZ ----AAA---G-------T---------A-GCC--------G-A-C---T--T--------A-AAG-T--  1523

CONSENSUS-A   AGAtAGggtTATAGAAaTAGgACAaagAattggTAGaGCtaTTctcaACATACCtagAAGAATCAGaCAg  2407
CONSENSUS-B   ----------------g---t-------gc-t?----------c---------?------a------  2462
CONSENSUS-C   -------a--------t--at-------'---t----------ctg----------------a------  2431
CONSENSUS-D   ----------------?---t-------gc-t?-------g----Tc-------C-c-c----A------  2426
CONSENSUS-E   ----------------G---C----g--gc-T-G-----c------C----------------------  2539
CONSENSUS-F   ------A---------GCTTTG------gc---------------t----------------A------  2462
CONSENSUS-G   ----------------g---t-------gC-T-------------------------g--A-----A  2485
CONSENSUS-O   t--Cg-CA-a--Ctt-GGgaT---------A--a----Ga----?T-----c--A--------T------  2412
CONSENSUS-U   ----?-???-------?-???----------?----?---?-???---?---------------?------  2263
CONSENSUS-A   ----------------------------------------------------------------------  2407
CONSENSUS-CPZ ------AA-A------GCTTTT---GTT-CACT----AT----AGA-----C-----GC-C--A-----A  1593

CONSENSUS-A   GGctt?GAAAggGctTTgctATAA                                              2430
CONSENSUS-B   -----g----------------                                                2486
CONSENSUS-C   -----t---gCA-------a----                                              2455
CONSENSUS-D   -----g----------a------                                               2450
CONSENSUS-E   -----A----------------                                                2563
CONSENSUS-F   -----g----------------                                                2486
CONSENSUS-G   -----a-----A----------                                                2509
CONSENSUS-O   -----a---?-Aa??--at-g---                                              2433
CONSENSUS-U   --?------???-----?-??---                                              2279
CONSENSUS-A   ----------------------                                                2430
CONSENSUS-CPZ --GC-G-----A-----A--T                                                 1614
```

HMMER Sequences in the Nef Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_SF2	HIVSF2	K02007	Sanchez-Pescador,R.	Science 227, 484 (1985)
D_ELI	HIVELI	K03454	Alizon,M.	Cell 46, 63 (1986)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68, 1586 (1994)
O_MVP5180	HIVMVP5180	L20571	Gurtler,L.G.	JVI 68, 1581 (1994)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
A_ROD	HIV2ROD	M15390	Clavel,F.	Nature 324, 691 (1986)
B_EHOA	HIV2EHOA	U27200	Rey-Cuille,M.A.	Virology 202, 471 (1994)
SD_MM251	SIVMM251	M19499	Franchini,G.	Nature 328, 539 (1994)
STM_STM	SIVSTM	M83293	Novembre,F.J.	Virology 186, 783 (1992)
VER_AGM3	SIVAGM3	M30931	Baier,M.	Virology 176, 216 (1990)
GRI_AGM677	SIVAGM677	M66437	Fomsgaard,A.	Virology 182, 397 (1991)
SAB_SAB1C	SIVSAB1C	U04005	Jin,M.J.	EMBO J. 13, 2935 (1994)
SYK_SYK	SIVSYK	L06042	Hirsch,V.M.	JVI 67, 1517 (1993)

The following alignment was generatede using the HMMER program as described in the introduction to this Part and in Part III. For simplicity, only representative types and subtypes are shown. An ordinary consensus sequence (lowercase signifies majority, uppercase signifies 50% or greater) was created from these sequences using MASE; this is not a "most likely sequence" based on a HMMER model (Part II). Annotation is based on HIV-1s, therefore the user should be cautious about its applicability to other PIV sequences.

```
NEF_CONSENSUS ATGGGtg?ca???g?tcaAa????????????????g?a?g???g?gggatgg???g??cT??ggga?a     34
A_U455        -------G--AGT-G-----A...........AAGA-C-GA...-T--A----CCT-AGG-TA--A-A-     55
B_SF2         -------G--AGT-G-----A...........CGTA-T-T-...-GT------TCT-CTA-AA----A-     55
D_ELI         -------G--AAT-G-----A...........AGTA-T-TA...-T-------CCT-CTA-AA----A-     55
O_ANT70C      -----AAA-GCATTGAG---A.............G-T-AATTT-A-------GCA-CAG-AA-A--A-     55
O_MVP5180     -----GAATGCAT-GAGC--A.............A-C-AATTT-CA------TCA-AAG-AA-A--T-     55
CPZGAB        -----AAC--AAT-G--T--A...........AGTA-TCT-...-TA------CCT-AGG-CA-AAGA-     55
A_ROD         -------CG-GTG-A--C--G.........AAGCATTCC-G-CCGCC-C--G--.......--ACAA--G-     55
B_EHOA        -----ATCAGCTG-T--C--G.........AAGCAATCC-A-CAGCA-CC-G--.......--GC----G-     55
SD_MM251      -------GAGCTATT--C-TG.........AGGCGGTCC-A-CCG-CT---GAT.......--GC-AC-G-     55
STM_STM       -------CATCTG-A--C--G.........AAGCAGC-C-A-CAGCAT---GA-.......--GC----G-     55
VER_AGM3      -----CTTGGGGAAC-----GCCGCAGCACAAGAAGCAGTT-...TC-CTT---CAC-CCT-GCAC....     63
GRI_AGM677    -----CTC--GCAAC--C--G...........AGGCAGCAA...CA---C-T-CTCAAG--CT--...C     52
SAB_SAB1C     -------G--AGA-C-----G.........CAGCAGCAGCGTCACTCACTG---.......--GT--TCC-     55
SYK_SYK       ------TCA-CGA-C-----A.......................TC-CAGCA-.......--TC-CTCGG     40

NEF_CONSENSUS ga?tgagacgagcgc??gca????c???????????????t?tg????????gcagatggggtggga??     70
A_U455        --A-----GA-A-T-CT---.....................................GCA---A-A--A--A---..     93
B_SF2         --A----------TGAGC--...CGA.............GC--AGCCAGCA--------------..    105
D_ELI         --A-A---A--A-TAATC--.....................................GCA-----------A---..     93
O_ANT70C      --A-----A--A-TAGAA-TTTCC-TGAG...........-C--AACCATGC---CC---A--A---..    111
O_MVP5180     --A--------T-CTCCT-TGATC-T.............CAACAACCATGT---CC---A--A---..    108
CPZGAB        --A-A--GGA---T-CAA--.....................................GCA-----G--A--A---..     93
A_ROD         --C-CTTG--G----GT---GGGG-T.............-G--GAGGGTATTGGA-C-AATC----GG    110
B_EHOA        --C-CTTG-------GGCG-GGAC-TCGT..........GGG-AGTCCTCA-G---GC--CA--AG..    111
SD_MM251      A-C-CTTG--G----GT-G-GAGA-T.............-A--GGAGACTCTT--GA-A-----A-GA    110
STM_STM       --C-CTTG-------GG-G-GAGA-T.............-A--GGAAACTCTTG--A---T-----GA    110
VER_AGM3      ........AA-A---GG--TACCAGA.............-A--GCTTGCTT------CC-C-TAT-GG    110
GRI_AGM677    --GG-CTG----G-AAGC-TGGGG-AGAC..........-GG-TGCTATTGT-C---CC-C-TATCGG    110
SAB_SAB1C     AGC--C---A-----CA-TGATTCAG.............-AC-ACATGCTT------CCAT--TT-GG    110
SYK_SYK       A-GG--A-TAC---ATT-GGTGGCGGCTATTTGGCAAGCAA-A-ACGCCGCTAC----C-A-T--TCG..    108

NEF_CONSENSUS ?g?atcatcgc?a???gacctaGaag?at??Ggcaag????????????????????????????g?c??g    101
A_U455        .-C-GT---T-A-...--TT----TAA--AT--AGCA.....................-T-ACA    132
B_SF2         .-C-GT---T-G-...-----G---AA-CAT--AGCA.....................AT-ACA    144
D_ELI         .-C-GT---T-G-...-----G---AA-CAT--GGCA.....................AT-ACA    132
O_ANT70C      .CAGATC--CAGG...--AT---C--CTAGA--AGG-.....................ATACCA    150
O_MVP5180     .-CTGTC--CAGG...--GT---C-ACTAGA--GGGA.....................ATATCA    147
CPZGAB        .-A-GTT---AAG...---------AG-CAC--AGCT.....................AT-ACT    132
A_ROD         G-A--AC----GG...TT--A-----G--CA-A--G-.....................-AGCA-    150
B_EHOA        .AG-----T--A-...T---C--G--G--CA-A----.....................-GGCT-    150
SD_MM251      T-G---C----A-...TC-----G--G--TA------.....................-G-TT-    150
STM_STM       A-G---GGG--C-...TC--A--GC-C--CA-A----.....................-G-TT-    150
VER_AGM3      GCA------AACT...CT--A-----AG-GC-A----.....................-C-TT-    150
GRI_AGM677    GCAG-----AAC-...-T--A-----AG-GC------.....................-C-TT-    150
SAB_SAB1C     GCAG-------AC...AT--A-----A--GC-C----AGCTTGAGAGATGGTTTAATTAGGCAA-GAGAC    177
SYK_SYK       .AGGC-G-T--AGCCTTG--GC-G--GC-TC-A----.....................-CTTG-    150

NEF_CONSENSUS ag?tc??????????????????????????????????????????aatactcca??????a?????g    116
A_U455        --CAGT.....................................-----AT--...TCT-CTAAT-    157
B_SF2         --TAGC.....................................-----AG--...GCT-CTAAT-    169
D_ELI         --TAGC.....................................-----AG--...AGT-CTAAT-    157
O_ANT70C      --T--C.....................................C-------T...CAA-ACAAT-    175
O_MVP5180     --T--C.....................................C-C-----T...CAA-ACAAT-    172
CPZGAB        --TAGG.....................................--C--C---...GAG-CTAATC    157
A_ROD         -AA--GCCCTCCTGTGAGGGACGGCAGTATCAGCAGGGAGACTTTATG---------TGGAAGGACCCA-    220
B_EHOA        -AC--GCCCTCTTGTGATGGACAA.....................--G---TTG.............    183
SD_MM251      --C--ACGCTCTTGTGAGGGACAGAAATACAATCAGGGGCAGTATATG---------TGGAGA-ACCCA-    220
STM_STM       -AC--ACACTCTTGTGAGCCTCAGAGATATAATGAAGGTCAATTTATG-----C--TTGGAAA-ACCCG-    220
VER_AGM3      -AGGAATCCTTAATTAGGAAG..............AGAAATGGTAAA-TG------.............    192
GRI_AGM677    -AAAAG.....................TCCTGGGGTAAAGGTAAA-TG------.............    183
SAB_SAB1C     TCC--T.....................................-GA--AGA-.............    192
SYK_SYK       --A--TACCTTAACTGAGCCC.................ATTGACCCAC-CGGG--G.............    189
```

```
NEF_CONSENSUS c?ga?ggagc?t??cta??????gaag?aca????????????????????????????????gA????ga   139
A_U455        -CAGTT-T--C-GG--G.....----CG--A.............................--G.....   186
B_SF2         -T--TT-T--C-GG---.....----C---A.............................--G.....   198
D_ELI         -T--CT-T--C-GG---.....----C---A.............................--A...--   188
O_ANT70C      -A-CCCTT--A-TC---.....---AGT--C.............................C-A.....   204
O_MVP5180     -A-CCCTT--A-TC---.....--CAGC--C.............................A-A.....   201
CPZGAB        AAACTCT---T-GG---.....--G-A-ATG.............................--C.....   186
A_ROD         -A-CA-A-AGGGAGAA-.....A-TTTGT-CAGGCAACAAAATATGGATGATGTAGATTCA--T...--   281
B_EHOA        .....-----AGAAGGGGGA...-G-AA---A...........GATTCAGATGAGGATGAT--G.....   228
SD_MM251      -T--A-A-AAAGAAAA-.....TT--C-T-CAGAAAACAAAATATGGATGATATAGATGAG--A...--   281
STM_STM       -A-CA-A-AGCGCTAAG.....CT--A-T-TAGACAGCAAAACATGGATGATGTAGATGAG--A...--   281
VER_AGM3      ..--A---AGAAAAT--CAA...----G-G-T...........AAATGGGATGAATGGTCT--T...--   242
GRI_AGM677    ..--C--CCGCCGC--GCAA...----G-G-C...........ACCTTTGATGAGTGGGAT--T...--   233
SAB_SAB1C     ..--A--T-TAAAAA-G.....A--CAT--AGGAAGACAGCCC.....TCGTGGTATGAT--G...--   245
SYK_SYK       ..--CA---AC-GGGG-CATTCG-G--G---AAAATTCTCACCTGGAGATATAGTGCAGGAT--AGGA--   257

NEF_CONSENSUS ?ga?ga?gaaGTagG?tT?ccaGTaagaCC?caagt?CC?ctaaGaccaaTgAC?tatAaattaGCa?Ta   199
A_U455        .--A-GA--C-----C--C-----T--G--A--G--A--TT----------T-----GGC---TT-T   255
B_SF2         .--G--A--G--G--T--T-----C-----T--G--A--TT----------T--C--GGC---TT--   267
D_ELI         GAGC--C--G--G--C--T-----C-----C--G--A--TT----------T--C---GA---TC--   258
O_ANT70C      .--G--A--------T--T------GC---T-----G--T-----G--------C------GG----T-T   273
O_MVP5180     .--T--G--T-----C--C----------T-----G--T-----G--------C-T----GC---CT-T   270
CPZGAB        .A-T--A--------A--C----------A--G--T--AACG----------T------GC---TT-T   255
A_ROD         T--T--CC-----A-AG-TT-T--C-C---AA----A--A----------AC---G---G---A--   351
B_EHOA        .--CA-T--------GG-C-GT--------CGGG--C--A--G--G--------A-TC---C-----G--   297
SD_MM251      T--T--CTTG-----GG-AT----G--G--AA----T--C------G-------T--C-----G---A--   351
STM_STM       T--TA-TCT------AG-AG-----CAT--AAG---C--A--T--GGA------T--C--G--G---A--   351
VER_AGM3      A--A--T--------A--T----------AAG---G--G-------A------C----------GG-G   312
GRI_AGM677    T--A--A--------C--C--T--GCA---T-G---C--CT------AG-----C----------G-G   303
SAB_SAB1C     T--A--A--------C--T--------G--TTGCT-A--A---C--G-------A-----G--G---A--   315
SYK_SYK       CACA-GTCTC-----T--C--T--GT-T--T---ACA--A--C---A--T-A--C--C-----G---A--   327

                                        3' LTR start ->
NEF_CONSENSUS GAccT?tcccacTTttTAAAAgaaAaGGGgGgACTGgAaGGgaTattttactctgaaAgaAGacaagaaA   268
A_U455        --T--CAG-TT-------------------------T---T--A--C----AC-G-A----------   325
B_SF2         --TA-TAG----------------------------C--A---GG--CC-----------G-      337
D_ELI         --T--CAG----------------------------C--A---GG--CA---AG--------G-      328
O_ANT70C      -----CAG-TT----------A------------T--A-------CC-T-A----GC-----      343
O_MVP5180     -----CAG-TT----------A--------T---T--A-------CC-T-AG---GC-----      340
CPZGAB        --T--T--A------------------------T--G-------CAGG----------G-      325
A_ROD         --TA-G--A--T--AA-----AC--G----------G------AG----------TA---      421
B_EHOA        ---A-G--T--T----------A------------T--C--TAG---G-----G--TA---      367
SD_MM251      --TA-G--T--T---A-------------------T-A----AG--C---------TAG--      421
STM_STM       -----G--A--T---A-----AGC-----------T-A-----AG---G--------TAG--      421
VER_AGM3      ---T-T--G--------------------------T-----T-A------C--C--G--GA-TC-G-      382
GRI_AGM677    ---T-T----------TC---------------T-------A----------------G--A-G-      373
SAB_SAB1C     ---T-TGG---------------------------T-A----AG------G---A-GA-G-      385
SYK_SYK       -----C--T--T---A-----A-C-----A------C-G--C--GAA---T-G-----AG---G-T----      397

NEF_CONSENSUS TccT?gAtct?Ta?gt?ta?aAcgaacaaGG?atc?T?cctGattGGcagaatTacaCaccaGGgCCaGG   330
A_U455        ----T-----G-GG--C--TC--AC------AT--T-C-----------------------------   395
B_SF2         ----T-----G-GGA-C--CC--AC------CTA-T-C-----------------------------   407
D_ELI         ----T-----T-GG--C--C---AC------C---T-C-----------A--C--------------   398
O_ANT70C      ----G-----T-GG--G--T---ACT--G--AT--T-C-----------C-----------A-----   413
O_MVP5180     ----G-----C-GGA-A--TC--ACT--G--AT--T-C-----------TG---------G--A-----   410
CPZGAB        ----T--C--C-GG--C--TC--AC------CT--T-C-----C---------C------A----A-----   395
A_ROD         --T-AA--A-A--CT-AG-A--G---G----G--AA-TG-A----------C-----T-AT--------   491
B_EHOA        -A--A--CACA--CT-AG-A--T---G----C--TG-GT---GA---------C-------AT--------   437
SD_MM251      --T-A--CA-G--CT-AG-A--G---G----C---A-A--A---------G-------CT----A-----   491
STM_STM       -AT-G--CA-G--CT-AG-A--A---G----G--AG-T--A----------------G----A-----   491
VER_AGM3      ----AA-C--G--C-CCCTC--T--GTGG--A---A-TGA-------A-TGC--GGT--GA---A-----   452
GRI_AGM677    ----GA--T-G--T-CC-TG-----GTGG--A--AA-AGA---------AGC----T-------C--G--   443
SAB_SAB1C     --T-G-----C--T-CTCTT--T--GTGG--C--AG-AGA--GA----------T---GAT--------   455
SYK_SYK       -T--TC----C--CC-AC-G--T--G--T--A---A-A...--CA--ATC--------CT----------   464

NEF_CONSENSUS aatcaGaTacCCaatga?cTTtGGaTggt?cTt?aAacTaGTaCCagt?ga??ta?aa??????gaagaa   387
A_U455        -------------C-A-CA----------G--AC--G----------T--TCC-GCTGAAGTA--G---   465
B_SF2         G---------T---C---C-----------G---C--G----------T--GCC-G-GAAGGTA-----G   477
D_ELI         G---------T---C-A-C-----------G--ACG-G----------T--TCC-C-GGAGGTA------   468
O_ANT70C      --C---G-T----C---CA----------TG--T--------------GTCAGA-G--GAGGCA---AG-   483
O_MVP5180     -CCT----T----C---CA----------TG--T-----G--------GTCAGC-G--GAGGCA--GAG-   480
CPZGAB        --CA----T----C-CTGT----------G---C--------G--CC-GACAGAGG-GCAGGTA---C--   465
A_ROD         -G-A-----------TT------G---CTA-GG--G----------A--TG-CCC-......C-----   555
B_EHOA        -G-A-----T--C-A-TT------C---CTA-GG--G-----G---A-AA-TA-GAT-.......-C----   501
SD_MM251      ---T----------A--CA-----C---CTA-GG---T----C--T--AA-TG--TC-........--T--G   555
STM_STM       G--A----------A-CAG----------CTA-GG--------------A--TA-GTC-.......A-T--G   555
VER_AGM3      -------------GATG---C--C-TC-G---T---T-G--------A--CC-GC-T.......--G---   516
GRI_AGM677    G--A--G-----G-GAGT------C-TC-G---T--G-----C-----G--CC-GC-T.......--G--G   507
SAB_SAB1C     T-C---G-------A-TG----------G---T-----G-----T--G--CC-GTC-........--G--G   519
SYK_SYK       --C-C-G-----GT-A-T----------TA-GGG-------T---AAT--GA--G-G.......-G-TGC   528
```

```
NEF_CONSENSUS gc????aagga?ggt???Gaga??cactg?cTg?TgCAtCCa???gc?cag?atgg??????gGatgacc  434
A_U455        --T...-CT-GA--A...----ACA--A-T--AC-A--C--TATATGC--AC----A...GTA-----TG  526
B_SF2         --C...--T--A--A...----ACA--A-CT--T-A--C--TATGAGC-T-C----G...AT---G---G  538
D_ELI         -AC...-CT--A--A...----CCA--A-CT--T-A--C--TATATGC---C----A...AT---G----  529
O_ANT70C      CTAGGA--TACAT--...----GGGCTAAT--CC--------GCATGTGCCC----A...TTT--A--TA  547
O_MVP5180     CTGGGT--TACAAA-...--AGATGCTA-T--TC-A------GCTTGTA-TC----A...GCT--G--TG  544
CPZGAB        --C...--T--A--A...--T-ACA----C---T-------CATATGT---C----G...AT---A--TG  526
A_ROD         -GG...G----CAC-...----CT-----CT-AG-A------...--A--AACAA-CAAGTTT-------  616
B_EHOA        C-A...G----C-AG...--A-CC--T--T---G--------...--A---ACCTCCGCATG--------  562
SD_MM251      --A...C----G-A-...----GG--T-ATT-AA----G---...--T--AAC-TCCAAGTG--------  616
STM_STM       --A...C----G-A-GAT-G--CA--T-AT---G--------...--A---ACACATCAGTG--------  619
VER_AGM3      --A...G--ACTT--...----GA--T--C---G--------...--A--AGTGA-G.....--A----  571
GRI_AGM677    --A...CGCA-CT--...----GA-----T---A--------...--A---ATG--G.....--A--T-  562
SAB_SAB1C     --T...---A-TA--...----AC-----C---C-------T...--T--AGTG-CCTATGA-------G  580
SYK_SYK       TTG...G--T-T-AG...---CATACACTC--CC-C------GCCT-T%%-G-CA-G...GGA-T-CGAG  587

NEF_CONSENSUS c?????????????agggagA?gt?cT?gt?TGGaagTTtgAcccca??cTaGc?tt?aa??Atgt?gc  479
A_U455        AA...........G--AA---A--GT-AA-G----------AGT-CC--G--A--A--AC-CAGA--  584
B_SF2         -G...........G--AA---A--GT-A--G----G-------AG--AA-----A--TC-TC-CA-G--  596
D_ELI         -G...........G--A--C-A--GT-AAAA----GA---A--AG--GA-----A--TG-GC-CAAG--  587
O_ANT70C      -A...........C-TAA---AA-A--GA-G----------TAGATCT----GCAAC-CCC----T--  605
O_MVP5180     -A...........C-C--G--GA-A--AAAA---C-------TAGATCAT---GC--A-CAC--A-A--  602
CPZGAB        AG...........G-CAA---G--CT-G--C---CGC------AG--GG--G--C--A-GAC--A-T--  584
A_ROD         -G...........C-T--G--GACA--A--C---G-------T---TTG--G--T-AT-GTT-C-AG--  674
B_EHOA        -C...........C-C-A---GACC--T--C---C--------T--CTC-----A-ATG-CT----G--  620
SD_MM251      -T...........TG------G--T--A-CG----------T--A-CT-----C-AC-CTT---AG--  674
STM_STM       -T...........TG------A--A--G--T-----A-----T--ACTAT----CCAT-CTT---AG--  677
VER_AGM3      -TGATGGAATCAACC-T-----A--CT-G--C----------T----TGT----AG-CC-AT---ACC-  641
GRI_AGM677    -TGATGGAATAGATC-T-----A--CT-G--C----------G-AGT-G--GG-GG-GT-CCGCC-  632
SAB_SAB1C     -C...........TG-AA---GACC--G--G----------ACTA-----AG-GG-CT----T--  638
SYK_SYK       -A...........TG-...--GAAC--CA-G-----C--CA-----CAC--G--TA-AC-CC-G--T-G  642

NEF_CONSENSUS ?????????gcacccagag?a?tt?????ac?agtcaggc?tg?ca?AggaagAgg??aag?????????  518
A_U455        TTATGAGCT---T--G---TTC-ATAAAG--                                        615
B_SF2         CCGAGAGCT---T--G---T-C-ACAAAG--TGC                                     630
D_ELI         CCGAGAGAT---T--G---TTC-ACAAAA--                                        618
O_ANT70C      TATGATAACT---------CTC--CCAGA-GG-C                                     639
O_MVP5180     CCTGCAAAA----------CTC--CCCCA-G                                        633
CPZGAB        CAGAGAACAA--T--G---T-C-ACAAAG--                                        615
A_ROD         TTTTATTCG-T--------G-A--TGGGC--A--------C--C--G--------TGG---.........  735
B_EHOA        CTTCAGCAG-TT-------G-G--TGGGT-TC-------AA--C--G--A-----TGG---.........  681
SD_MM251      ATATGCTAGAT-------AG-G%-TGGAAG-A--------C--T--G---------TT-GAAGAAGG    740
STM_STM       CTTTGTTAGA--------AG-G--TGGTTCTA--------T--C--A---------TTG--AGAAGGCTA  747
VER_AGM3      TAAC.........AG-A-AT-TC-CACTG--AT-CAT-ATC-TGGCA--AGG                    684
GRI_AGM677    GGAC.............ATG--TAAGG--.............ATGC-C---C-T-CA---CGC         669
SAB_SAB1C     TTGGAGGCT----------C-GG-GCCGTCAGCACA----                               678
SYK_SYK       GAGA.........................TG-C-C-CTTGCAGTT--AG---CAAAC--.........  676

NEF_CONSENSUS ????????????gctaga?tgaa?gcaagaggaAtacca?c?agt                          547
A_ROD         ............--G---C----A---------------TTT---                          768
B_EHOA        ............----A-C---GA--------------TA-AGAG                          714
STM_STM       ACCGCAAGAGGCCT-CT-AA--TG--TGACAAG-AGGA-A-A--C                          792
SYK_SYK       GAAAACCACAAGAGCT-CAGTC-GCACTT--T--G-A--T-TCC-                          721
```

```
CONSENSUS-A ATGGGTGGCAAGTGGTCAAAAAg?AGCAtAGTGGgATGGCCTgAGGTTAggaAAAGAATgAGAcAAACt.    68
CONSENSUS-B ----------?-t--t?-g-?t---------act--A---g-----------g-g--?              66
CONSENSUS-C -----G----------T--------T----------CT--A--AG-------A---AG----.        69
CONSENSUS-D ----------A----------T--T-----t----------CTA-A---G-------A---Ag----?    69
CONSENSUS-E -----A----------T----------C----c---G-------A-aG------.                 69
CONSENSUS-F ------?----------T--T-----T--?-------CTA-A---G-----?-?----??--C.        63
CONSENSUS-G -----A----------T--T-----T----------CA--A---G-------A---AG----C         70
CONSENSUS-O -----?AA?GCA-?-AG?---?---AAT-T-??------?-A-?A--A--AG-?---------?G-?-?. 55
CONSENSUS-U ------??----------???-?????????---------????-?---G-------A--?-?---??  46

CONSENSUS-A .....................................???CCTaCAGCAGCAAaAGGAGTAGGAGC      94
CONSENSUS-B ????????????????????????????????????????---GAgC--------g-t--g--g-----  92
CONSENSUS-C .....................................-AGC--------G-G----------          95
CONSENSUS-D ????????????????.....................gA-C--------G-T--G---------       95
CONSENSUS-E .....................................---c-----a--g----------            95
CONSENSUS-F .....................................---C--------G-?--G--G-----          88
CONSENSUS-G CACCAGCAGCAGAAAGAAAA.....................GAAG--------G----------      116
CONSENSUS-O .....................????CT???CCT?A?---?AAC--TG?---CCT---------??      83
CONSENSUS-U ????????.....................................--?C--?---?-??----------  67

CONSENSUS-A AGtATCTCAAGAT.............TTAGATAaAcATGGAGCAaTCACAAGCAGTAATaca?ca??a    146
CONSENSUS-B --------g---c.............c-g--a--------------------t--c------g--gct    147
CONSENSUS-C --CG--------C.....................T--------C-T--------C------C-CACC    150
CONSENSUS-D --------G---C.............C-G--A--------G----------T--C------g--agT    150
CONSENSUS-E ----------.............C-----------------g-A-----T-------Tg......      144
CONSENSUS-F ---G--------C.............-----A-G--GG--G-----T----?---C-----TAG-G-T    141
CONSENSUS-G ---------G---.............----C--GG---------G-T--------C------G--GC-    171
CONSENSUS-O ??-C--CAGG--?.....................----CA?CTAGA--?-G?--A?----TTCCC-?--TC-TCA-  130
CONSENSUS-U --?----------??????????????----?---???-------??-?-------C-?-?--G---CT  111

CONSENSUS-A AcTaAccCtagTTGtgCCTGGCTGGAA???GCacAAGag?.....GA?......Gagga?...GTAGGCT  196
CONSENSUS-B --c--tg--Ga----------a---...----------??...--g??????-----g???--g--t-  199
CONSENSUS-C T-A--TG--GA----------C--...--G------.....--GGAA...-GA--A...-------    205
CONSENSUS-D -----tG--gaC----------a---...---------a.....--GAGC...-----G...--G----  205
CONSENSUS-E -A---Tg--GA-----T-------ag-...----------.....--a??????----gG...-------  196
CONSENSUS-F -A---T---GAC-TG----------?...----------.....--A.....--?--A...-------  191
CONSENSUS-G -AC--T---GA----------...----------.....--GGAG...TCA--A...-------      226
CONSENSUS-O -AC--TG-AGCCCT---A-TC--A--?...AG?--C?-A.....---.....--?-----...-----?-  175
CONSENSUS-U -A---T?--??----?????????---...?---?---?.....---???...-----G...-------  147

CONSENSUS-A TtCCAGTcAGgCCaCAGGTaCCTtTaAGACCAATGACTTAtAAGGgAGCTgT?GATCTCAGCcaCTTTTT  265
CONSENSUS-B ----------a--t----------c----c----?-a-----t----------                  268
CONSENSUS-C ----------A--T-----G----------AT-T---------TT-----                     274
CONSENSUS-D ----------A--t----------C--A-a------a-----t----------                  275
CONSENSUS-E ----------g---------c----------T-T-----T---TT------                    266
CONSENSUS-F ----------A--T----------?--?----------?----------T---------            257
CONSENSUS-G ----------A----------C-G----------T---AA-T---T-T---------TT-----       295
CONSENSUS-O -?-----A??A--T--A--G---C----G--------C-?---A-?---?T-T--C------TT------  239
CONSENSUS-U --------?-?--?---------?-?----------?-?---?--?---T-T-----?---??------  205

                  3' LTR ->
CONSENSUS-A AAAAGAAAAGGGGGGACTGGATGGGTTAATTtACTCcaaGAaAAGACAAGAAATCCTTGATCTGTGGGTC  335
CONSENSUS-B ----------a---c----------C-a----------t----------                     338
CONSENSUS-D ----------A---C------GG-----a--g--------G----------T------            345
CONSENSUS-E ----------c----------g---------g--------cT-A------                     336
CONSENSUS-F ----------A----------?-------?---------G----------?                    324
CONSENSUS-O ----------A--------?----------C-T--?---GC---------G-----?---?-?        304
CONSENSUS-U ----------?--?---??---???----------?---------                          267

CONSENSUS-A TAtaACACACAAGGaTtCTTCCCtGATTGGCAGAAtTACACACCAGGGCCAGGgAccAGATTcC...CCAC  402
CONSENSUS-B --cc----------c-a----------c----------t-----at?..-----                 405
CONSENSUS-D --C----------CA----------C----------T-----At...-----                   412
CONSENSUS-E -----t--------c----------?--c----------t-----a-...-----                402
CONSENSUS-F --CC--                                                                  330
CONSENSUS-O ---?----T--G----------???--------?--A-----A?-?--?---...-----          363
CONSENSUS-U --?C-?--------C-?----------?--?----------??----?-...-----             326

CONSENSUS-A TaACaTTTGGaTGGTGCTtCAAgCTAGTACCAgTtGATCCAGc?GAaGTAGAG...gAaGccActG?aGG  467
CONSENSUS-B -G--c----------g----aga-g-----a...--g----a--a---                       472
CONSENSUS-D -g--C----------g----------CaG--G-----A...--g-------A---                479
CONSENSUS-E -GTGT----------c---aga---------...--G-A--acaA---                       469
CONSENSUS-O -G----------TG--T--A--?--------GTCAG?--AA--G-C---?AGACT?-G?-A-AC-??   426
CONSENSUS-U -G???--?----------?---?--------?-???----??----------...--?----A-?---- 377

                     <-              NFAT-1              ->
CONSENSUS-A AGAGAaCAACAGtcTAtTACACCCtATaTGcCAACATGGAaTgGAtGAtGA?gA?AgAgAAgtgtTaATg  535
CONSENSUS-B ----------ct-g----------ga---tg-----g--------CcCg--G----------ga-      542
CONSENSUS-D -----c----t-CT-G----------g-----G----------G--CcCG--G---c---------a-  549
CONSENSUS-E ---a------T-C--g--------C--gA----G--------a--g--c--a--a---------C-G---  539
CONSENSUS-O T--????GCT-?---?C-?--T--AGC?--T???------??T--?---?CAC--???--?A-AC-?-?? 473
CONSENSUS-U ----------T-?--G--------?--?A----?----------?--?-?---A---------?---?? 436
```

HIV1 NEF CONSENSUS

```
                                        <- IL2/NRE ->
                                     <-    NRF     ->
CONSENSUS-A TGGAagTTTGACAGtagcCTGGCATTaAaACACAgAGCt?aaGAgcTGCATCCGGAGTtCTAc...AAAG     601
CONSENSUS-B ----g----------cc----a-----tc-t----tg--ccg-----------------a----...--g-     609
CONSENSUS-D ----GA---A-----C--A--A-----TG-g----AG--CCG----a----------------...----     616
CONSENSUS-E ---------------gC---A---CG--------t---CcG---A--------a----a---T...----     606
CONSENSUS-O ---?-------T--ATC??-A-GC???-C?--T?T?--?-TG??AA??--C--A---C---T-C??--G?     525
CONSENSUS-U --------------C??---A---C?--?----??---C-G---????----------?----...----     491

CONSENSUS-A ACTGC---TGA                                                               609
CONSENSUS-B -----?--TGACATCGATGTGTCTACAAGGGACTTTCCGCTGGGGCATTTCCAGGGAGGCGCGGCTGGGC     677
CONSENSUS-D -----                                                                     621
CONSENSUS-E --------                                                                  614
CONSENSUS-O -?                                                                        526
CONSENSUS-U -----                                                                     496

CONSENSUS-B GGGACTGGGGAGTGGCGAGCCCTCAGATGCTGCATA                                      712
```

```
LOCUS       HIV93TH253        9720 bp    DNA             VRL       18-NOV-1996
DEFINITION  HIV-1 isolate 93th253 from Thailand, complete genome.
ACCESSION   U51189
KEYWORDS
SOURCE      Human immunodeficiency virus type 1.
  ORGANISM  Human immunodeficiency virus type 1 (HIV-1)
            Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;
            Retroviridae; Lentivirinae.
REFERENCE   1  (bases 1 to 9720)
  AUTHORS   Gao,F., Robertson,D.L., Morrison,S.G., Hui,H., Craig,S.,
            Fultz,P.N., Decker,J., Girard,M., Shaw,G.M., Hahn,B.H. and
            Sharp,P.M.
  TITLE     The heterosexual HIV-1 epidemic in Thailand is caused by an
            intersubtype (A/E) recombinant of African origin
  JOURNAL   J. Virology 70 (10), 7013-7029 (1996)
REFERENCE   2  (bases 1 to 9720)
  AUTHORS   Foley, Brian T.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-1996) HIV Database, LANL, Mail Stop K710, Los
            Alamos, NM 87545, USA
COMMENT     Subtype E.  93TH253 has been previously named as CMU010 or
            302053.  It is from a Thai patient with end-stage AIDS.  The
            isolate has been adapted to growth in H9 cells.
            There are two frameshifts in 93TH253.  There is a G insertion at
            position 5798 which shifts the vpr gene.  The second is in vpu,
            where an A deletion occurs right after the ATG start codon.
FEATURES             Location/Qualifiers
     source          1..9720
                     /organism="Human immunodeficiency virus type 1"
                     /strain="93th253"
                     /chromosome="complete genome"
     LTR             1..633
                     /note="5' LTR"
     misc_binding    363..372
                     /note="nf-kappa-b binding site"
     mRNA            455..9638
     repeat_region   455..551
     mRNA            455..9638
     misc_feature    763
                     /note="Description: major 5' splice donor site"
     CDS             807..2279
                     /gene="gag"
                     /note="Description: Gag polyprotein gene; cleaved into
                     Capsid, Matrix and other proteins"
                     /codon_start=1
                     /product="Gag precursor polyprotein"
                     /translation="MGARASVLSGGKLDAWEKIRLRPGGKKKYKMKHLVWASRELERF
                     ALNPGLLETAEGCQQLIEQLQSTLKTGSEELKSLYNTIATLWCVHQRIEVKDTKEALD
                     KIEEVQKKSQQKKQQAAADTGSSSKVSQNYPIVQNAQGQMVHQPLSPRTLNAWVKVIE
                     EKGFNPEVIPMFSALSEGATPQDLNMMLNIVGGHQAAMQMLKETINEEAAEWDRVHPV
                     HAGPIPPGQMREPRGSDIAGTTSTLQEQIGWMTNNPPIPVGDIYKRWIILGLNKIVRM
                     YQPVSILDIRQGPKEPFRDYVDRFYKVLRAEQATQEVKNWMTETLLVQNANPDCKSIL
                     KALGTGATLEEMMTACQGVGGPSHKARVLAEAMSQAQHANIMMQRGNFKGQTRIKCFN
                     CGKEGHLARNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSNKGRPGNFPQSK
                     PEPTAPPAENWGMGEEQKDKEHPPPSVSLKSLFGNDPLSQ"
     CDS             join(807..2099,2098..5083)
                     /note="Gag-Pol fusion is caused by ribosomal slip on
                     'tttttt' slippery;sequence at 2096-2101, potentiated by
                     stem-loop structure at 2112-2135; Polyprotein cleaved into
```

```
                    protease, integrase, RNAse, reverse transcriptase etc"
                    /codon_start=1
                    /product="Gag-Pol fusion polyprotein"
                    /translation="MGARASVLSGGKLDAWEKIRLRPGGKKKYKMKHLVWASRELERF
                    ALNPGLLETAEGCQQLIEQLQSTLKTGSEELKSLYNTIATLWCVHQRIEVKDTKEALD
                    KIEEVQKKSQQKKQQAAADTGSSSKVSQNYPIVQNAQGQMVHQPLSPRTLNAWVKVIE
                    EKGFNPEVIPMFSALSEGATPQDLNMMLNIVGGHQAAMQMLKETINEEAAEWDRVHPV
                    HAGPIPPGQMREPRGSDIAGTTSTLQEQIGWMTNNPPIPVGDIYKRWIILGLNKIVRM
                    YQPVSILDIRQGPKEPFRDYVDRFYKVLRAEQATQEVKNWMTETLLVQNANPDCKSIL
                    KALGTGATLEEMMTACQGVGGPSHKARVLAEAMSQAQHANIMMQRGNFKGQTRIKCFN
                    CGKEGHLARNCRAPRKKGCWKCGKEGHQMKDCTERQANFFFRENLAFQQREARKFSSE
                    QTRTNSPTSSRDLWDEGRDSLPSEAGAERQGPEPTFSFPQITLWQRPLVTVKIGGQLK
                    KALLDTGADDTVLEDINLPGKWKPKMIGGIGGFIKVKQYDQILIEICGKKAIGTVLVG
                    PTPVNIIGRNMLTLIGCTLNFPISPISTVPVKLKPGMDGPRIKQWPLTEEKIKALTEI
                    CADMEREGRISKIGPENPYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIP
                    HPAGLKKKKSVTVLDVGDAYFSVPLDESFRKYTAFTIPSTNNETPGIRYQYNVLPQGW
                    KGSPAIFQASMTKILEPFRSKNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELREHLLK
                    WGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIELPEKESWTVNDIQKLVGKLNWASQ
                    IYAGIKVKQLCKLLRGTKALTDIVTLTEEAELELAENREILKDPVHGAYYDPSKDLIA
                    EIQKQGQDQWIYQIYQEPFKNLKTGKYARKRSAHTNDVKQLAEVVQKVVMESIVIWGK
                    TPKFKLPIQKETWETWWMDYWQAT$IPEWEFVNTPPLVKLWYQLEKDPIAGAETFYVD
                    GAANRETKLGKAGYVTDRGRQKVVSLTETTNQKTELHAIHLALQDSGSEVNIVTDSQY
                    ALGIIQAQPDRSESELVNQIIEKLIEKDKVYLSWVPAHKGIGGNEQVDKLVSSGIRKV
                    LFLDGIDKAQEEHERYHSNWRAMASDFNLPPIVAKEIVASCDKCQLKGEAMHGQVDCS
                    PGIWQLDCTHLEGKVILVAVHVASGYVEAEVIPAETGQETAYFLLKLAGRWPVKVVHT
                    DNGSNFTSAAVKAACWWANVKQEFGIPYNPQSQGVVESMNKELKKIIGQVREQAEHLK
                    TAVQMAVFIHNFKRKGGIGGYSAGERIIDIIASDLQTKELQKQITKIQKFRVCYRDSR
                    DPIWKGPAKLLWKGEGAVVIQDNSDIKVVPRRKVKIIKDYGKQMAGDDCVAGRQDED"
stem_loop           2112..2135
                    /note="stem-loop potentiates ribosomal slip on tttttt
                    slippery sequence to produce Gag-Pol fusin polyprotein"
CDS                 5028..5606
                    /gene="vif"
                    /note="Description: Viral Infectivity factor"
                    /codon_start=1
                    /product="Vif protein"
                    /translation="MENRWQVMIVWQVDRMRIRTWNSLVKHHMYISKKAKQWFYRHHY
                    ESQHPKVSSEVHIPLGEARLVIRTYWGLQTGEKDWQLGHGVSIEWRQRKYSTQIDPDL
                    ADQLIHLQYFDCFSDSTIRRAILGQVVRRRCEYPSGHNKVGSLQYLALKALTTPKRIR
                    PPLPSVKKLTEDRWNKPQKIKGHRENPTMNGH"
misc_feature        5377
                    /note="3' splice acceptor site"
misc_feature        5451
                    /note="5' splice donor site"
CDS                 5546..5830
                    /gene="vpr"
                    /codon_start=1
                    /product="Vpr"
                    /note="G insertion at 5798 shifts frame for COOH-terminal
                    region of protein."
                    /translation="MEQAPEDQGSQREPYNEWTLELLEELKNEAVRHFPRPWLHGLGQ
                    YIYNNYGDTWEGVEAIIRILQQLLFVHFRIGCQHSRIGIFAREKRQEWSW"
CDS                 join(5818..6032,8363..8456)
                    /gene="tat"
                    /codon_start=1
                    /product="Tat"
                    /translation="MELVDPNLEPWNHPGSQPTTACSKCYCKKCCWHCQLCFLKKGLG
                    ISHGRKKRKHRRGTPQSRKDHQYPIPEQPLPIIRGGNPTDPKESKKEVASKAETDPCD"
CDS                 join(5957..6032,8363..8679)
```

```
                     /gene="rev"
                     /codon_start=1
                     /product="Rev"
                     /translation="MAGRSGSTDEELLRAVRIINILYQSNPYPSSEGGTRQTRKNRRR
                     RWRARQRQIRAISERILSTCLGRSTEPVPLQLPPLERLHLDCSEDCGTSGTQQSQGTE
                     TGVGRPQISGESSVILGPGTKN"
     intron          6033..8362
                      /note="Tat-Rev intron"
     CDS             join(6053..6057,6056..6297)
                     /gene="vpu"
                     /pseudo
                     /note="single base deletion, frameshift, near 6057 results
                     in premature termination"
                     /product="Vpu (defective)"
     CDS             6215..8803
                     /gene="env"
                     /pseudo
                     /note="premature stop codon at 8036-8038; tga here, tgg in
                     others"
                     /note="envelope glycoprotein, signal peptide, gp120 and
                     gp41"
                     /codon_start=1
                     /product="Envelope"
                     /translation="MRVKETQINWPNLWKWGTLILGLVIMCSASNNLWVTVYYGVPVW
                     RDADTTLFCASDAKAHETEVHNVWATHACIPTDPNPQEMHLENVTENFNMWKNNMVEQ
                     MQEDVISLWDQSLKPCVKLTPLCVTLNCTNANWTNANVTNVNNNVTNIVGNITEEVRN
                     CSFNMTTELIDKKQKVYALFYKLDIRQMNSNSSEYRLINCNTSVIKQACPKVSFDPIP
                     IHYCTPAGYAIIKCNDKNFNGTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAEEKIII
                     RSENLTNNAKTIIVHLHESVEINCTRPFYNKRTRTSIGQGRVLYRTGDITGNIGKPYC
                     EINGTKWNKVLNQVTEKLKEHFNNRNISFQPPSGGDLEITMHHFICRGEFFYCNTTRL
                     FNNTCIGNKTMKECNDTIILPCKIKQIINMWQGVGQAMYNPPISGNINCVSNITGILL
                     TRDGGGGNGTNNEETFRPGGGNMKDNWRNELYKYKVVEIEPLGIAPTKAKRRVVEREK
                     RAVGIGAMIFGFLGAAGSTMGAASITLTVQARQLLSGIVQQQSNLLRAIEAQQHLLQL
                     TVWGIKQLQARVLAVERYLKDQKFLGLWGCSGKIICTTAVP"
     CDS             8805..9419
                     /gene="nef"
                     /codon_start=1
                     /product="Nef"
                     /translation="MGGKWSKSSIVGWPQVRERIKQTPPAAEGVGAVSQDLDKHGAVT
                     SSNMNNADCVWLRAQEEEGVGFPVRPQVPLRPMTYKGAFDLSFFLKEKGGLEGLVYSK
                     KRQEILDLWVYHTQGFFPDWHNYTPGPGIRYPLCFGWCFKLVPVDPREVEEDNKGENN
                     CLLHPMSQHGIEDEEREVLIWKFDSALARRHIARELRPEFYKDC"
     LTR             9088..9720
                     /note="3' LTR"
     repeat_region   9542..9638
     polyA_signal    9614..9619
     polyA_site      9638
     misc_feature    6053..6297
                     /gene="vpu (defective)"
                     /pseudo
                     /note="Description: vpu has 1 bp deletion frameshift near
                     6061"
BASE COUNT     3482 a   1711 c   2362 g   2165 t
ORIGIN
         1 tggaagggct agtttactcc aagaaaagac aagagatcct tgacttatgg gtctatcata
   5' LTR U3 ->
        61 cacaaggctt cttccctgat tggcataact acacaccagg gccagggatc agatatccac
       121 tgtgttttgg atggtgcttc aaactagtac cagttgaccc aagagaagta gaggaggaca
       181 acaaaggaga aaacaactgc ctgttacacc ccatgagcca gcatggaata gaggacgaag
```

```
 241 aaagagaagt gctgatatgg aagtttgaca gtgccctagc acgaagacac atagcccgag
 301 aactgcgtcc agagttctat aaagactgct gacaaagaag tttctaacta ggacttccgc
                                    <- nef cds end
 361 tggggacttt ccaggggagg tgtggccggg gcggagttgg ggagtggtta accctcagat
       -------- --- Nf Kappa B
 421 gctgcataaa agcagccgct tttcgcttgt actgggtctc tcttgttagg ccaggtcgag
        ------ --                LTR U3/\LTR R repeat
     TATA box signal
 481 cccgggagct ctctggctag caggggaacc cactgcttaa agcctcaata aagcttgcct
                                                          ---- --
                                                      poly-A signal
 541 tgagtgctta aagtggtgtg tgcccatctg tgttaggact ctggtaacta gagatccctc
   LTR R repeat/\LTR U5
 601 agatcactct agactgagta aaaatctcta gcagtggcgc ccgaacaggg acttgaaagc
                                    <- LTR U5
 661 gaaagttaat agggactcga aagcgaaagt tccagagaag ttctctcgac gcaggactcg
 721 gcttgctgag gtgcacacag caagaggcga gagcggcgac tggtgagtac gccaaatttt
                                                  /\ major 5' sj
 781 gactagcgga ggctagaagg agagagatgg gtgcgagagc gtcagtatta agtgggggaa
                        gag cds ->
                    p17 MA start
 841 aattagatgc atgggaaaaa attcggttac ggccaggggg aaagaaaaaa tataaaatga
 901 aacatttagt atgggcaagc agagaattag aaagatttgc acttaaccct ggccttttag
 961 aaacagcaga aggatgtcaa caattaatag aacagttaca gtcaactctc aagacaggat
1021 cagaagaact taaatcatta tataatacaa tagcaaccct ctggtgcgta caccaaagga
1081 tagaggtaaa agacaccaag gaagctttag ataaaataga ggaagtacaa aagaaaagcc
1141 agcaaaagaa acagcaggca gcagctgaca caggaagcag cagcaaagtc agccaaaatt
1201 accctatagt gcaaaatgca caagggcaaa tggtacatca gcctttatca cctagaactt
gag p17 MA/\gag p24 CA
1261 tgaatgcatg ggtgaaagta atagaagaaa agggttttaa cccagaagta atacccatgt
1321 tctcagcatt atcagaggga gccaccccac aagatttaaa tatgatgcta aatatagtgg
1381 ggggacacca ggcagcaatg caaatgttaa aagaaaccat caatgaggaa gctgcagaat
1441 gggatagggt acacccagta catgcagggc ctattccacc aggtcagatg agggaaccaa
1501 ggggaagtga catagcagga actactagta cccttcaaga acaaatagga tggatgacaa
1561 acaatccacc tatcccagtg ggagacatct ataaaaggtg gataatcctg ggattaaata
1621 aaatagtaag aatgtatcaa cctgttagca ttttggacat aagacaaggg ccaaaagaac
1681 ccttcagaga ctatgtagat aggttctata aagttctcag agcggaacaa gccacacagg
1741 aggtaaaaaa ctggatgaca gaaaccttgc tagtccaaaa tgcgaatcca gactgtaagt
1801 ccattttaaa agcattagga acaggagcta cattagaaga aatgatgaca gcatgccagg
1861 gagtgggagg acctagccac aaggcaaggg ttttggctga ggcaatgagc caagcacaac
                                gag p24 CA/\gag p2 X
1921 atgcaaatat aatgatgcag agaggcaatt ttaagggcca gacaagaatt aagtgcttca
           gag p2 X/\gag p7NC
                    \gag-pol TF
1981 actgtggcaa agaaggacac ctagccagaa attgcagggc ccctagaaaa aagggttgtt
2041 ggaaatgtgg aaaggaagga catcaaatga aagactgcac tgagagacag gctaatttttt
                                                                  -----
                                                          gag-pol ribosomal
                                                          slip site
2101 tagggaaaat ttggccttcc aacaagggaa ggccagggaa ttttcctcag agcaaaccgg
     -           ^^^^^^^^^^       ^^^^ ^^^^^   gag p7/\gag p6
                    stem    loop    stem
2161 agccaacagc tccgccagca gaaaactggg ggatggggga ggagcagaaa gacaaggaac
2221 atcctcctcc ttcagtttcc ctcaaatcac tctttggcaa cgacccctg tcacaataaa
              gag-pol TF/\pol protease                    gag cds <-
                                                                  p6 end
2281 aataggagga caactgaaag aagctctatt agatacagga gcagatgata cagtattaga
2341 agatataaat ttgccaggga aatggaaacc aaaaatgata gggggaattg gaggttttat
2401 caaggtaagg caatatgatc agatacttat agaaatttgt ggaaaaaagg ctataggtac
```

```
2461 agtattagta ggacctacac ctgtcaacat aattggacga aatatgttga ctcagattgg
2521 ttgtactcta aatttcccaa ttagtcctat tgacactgta ccagtaaaat taaagccagg
        pol protease/\pol reverse transcriptase
2581 aatggatgga ccaaaggtta aacagtggcc attgacagaa gaaaaaataa aagcattaac
2641 agaaatttgt aaagagatgg aagaggaagg aaaaatctca aaaattggac ctgaaaatcc
2701 atacaatact ccagtatttg ctataaagaa aaaggacagc accaaatgga ggaaattagt
2761 agatttcaga gaactcaata aaagaactca ggacttttgg gaagttcaat taggaatacc
2821 gcatccagca ggtttaaaaa agaaaaaatc agtaacagtg ctagatgtgg gagatgcata
2881 tttttcagtt cctttagatg aaagctttag aaagtatact gcattcacca tacctagtat
2941 aaacaatgag acaccaggaa tcagatatca gtacaatgtg ctgccacagg gatggaaagg
3001 atcaccggca atattccaga gtagcatgac aaaaatctta gagcccttta gaataaaaaa
3061 tccagaaatg gttatctatc aatacatgga tgacttgtat gtaggatctg atttagaaat
3121 agggcagcac agaacaaaaa tagaggagct aagagctcat ctattgagct ggggatttac
3181 tacaccagac aaaaagcatc agaaggaacc tccattcctt tggatgggat atgaactcca
3241 tcctgacaga tggacagtcc agcctataga actgccagaa aaagacagct ggactgtcaa
3301 tgatatacag aaattagtgg gaaaactgaa ttgggcaagt caaatttatg cagggattaa
3361 ggtaaagcaa ctgtgtaaac tcctcagggg aactaaagca ctaacagata tagtaccact
3421 gactgaagaa gcagaattag agttggaaga gaacagggag attctaagaa tccctgtgca
3481 tggagtatat tatgacccat caaaagactt agtagcagaa gtacagaaac aaggggcagga
3541 ccaatggaca tatcaaattt atcaagagcc atttaaaaat ctaaaaacag gaaaatattc
3601 cagaaaaagg tctgctcaca ctaatgatgt aagacaatta acagaagtgg tacaaaaaat
3661 agccacagaa agcatagtaa tatggggaaa gacccctaaa tttagactac ccatacaaag
3721 agaaacatgg gaaacatggt ggatggagta ttggcaggct acctggattc ctgaatggga
3781 gtttgttaat acccctcctc tagtaaaatt atggtaccaa ttagaaaaag accccatagt
3841 gggagcagag actttctatg tagatggggc agctagtagg gagactaagc taggaaaagc
pol reverse transcriptase/\pol RNAse
3901 agggtatgtc actgacaggg gaagacaaaa ggtaatttcc ctaactgaga caacaaatca
3961 aaagactgaa ttacatgcga tccatttagc cttgcaggat tcaggatcag aagtaaatat
4021 agtaacagac tcacaatatg cattaggaat cattcaggca caaccagaca ggagtgaatc
4081 agaagtagtc agccaaataa tagaggagct aataaaaaag gaaaaagtct acctgtcatg
4141 ggtaccagca cacaagggga ttggaggaaa tgaacaagta gataaattag tcatttcagg
4201 aatcaggaag gtactatttt tagatgggat aaataaggct caagaagaac atgaaagata
            pol RNAse/\pol integrase
4261 tcacagcaat tggagaacaa tggctagtga ctttaatttg ccacctatag tagcaaagga
4321 aatagtagcc aactgtgata aatgtcaact aaaaggggaa gctatgcatg gacaagtaga
4381 ctgtagtcca gggatatggc aattagattg cacacatcta gaaggaaaag tcatcctggt
4441 agcagtccac gtggccagtg gatatataga agcagaagtt atcccagcag aaacaggaca
4501 ggagacagca tactttctgc taaaattagc aggaaggtgg ccagtaaaag taatacacac
4561 agacaacggt agcaatttca ccagcgctgc agttaaagca gcctgttggt gggccaatgt
4621 ccgacaggaa tttgggatcc cctacaatcc ccaaagtcaa ggagtagtag aatcaatgaa
4681 taaggaatta aagaaaatca tagggcaggt aagggagcaa gctgaacacc ttaagacagc
4741 agtacaaatg gcagtattca ttcacaattt taaaagaaaa ggggggattg gggggtacag
4801 tgcaggggaa agaataatag acataatagc aacagacata caaactaaag aattacaaaa
4861 acaaattaca aaaattcaaa attttcgggt ttattacagg gacagcagag acccaatttg
4921 gaaaggacca gcaaaactac tctggaaagg tgaaggggca gtagtaatac aagacaatag
4981 tgatataaaa gtagtaccaa gaagaaaagc aaagatcatt agggattatg gaaaacagat
                                          vif cds start ->
5041 ggcaggtgat gattgtgtgg caggtagaca ggatgaggat tagaacatgg aacagtttag
5101 taaaacacca catgtatatc tcaaagaaag ctaaacagtg gttttataga catcattatg
                                                        <- pol cds
                                                           integrase end
5161 aaagccagca tccaaaggtg agttcagaag tacatatccc actaggagaa gctagattag
5221 taataagaac atattggggt ctgcagacag gagaaaagga ctggcaattg ggtcatggag
5281 tctccataga atggaggcag agaaaatata gcacacaaat agatcctgac ctagcagacc
5341 aactgattca tctacaatac tttgactgtt tttcagactc taccataagg agagccatat
5401 taggacaagt agttagacgt aggtgtgaat atccatcagg acataacaag gtaggatccc
5461 tacaatattt ggcactgaaa gcattaacaa caccaaaaag gataaggcca cctctgccta
5521 gtgtaaagaa attaacagaa gatagatgga acaagcccca gaagatcaag ggtcacagag
                  vpr cds start ->
```

```
5581 agaaccctac aatgaatgga cattagaact gttagaggag cttaaaaatg aagctgttag
                              <- vif cds end
5641 acattttcct aggccctggc tccatggctt aggacagtac atctataaca attatgggga
5701 tacttgggaa ggggttgaag ctataataag aattttgcaa caactactgt ttgttcattt
5761 cagaattggg tgtcaacata gcagaatagg cattttgcc agggagaaga ggcaggaatg
                                                    tat cds start ->
5821 gagctggtag atcctaacct agagccctgg aatcatccgg gaagtcagcc tacaactgct
          <- vpr cds end
5881 tgtagcaagt gttactgtaa aaaatgttgc tggcattgcc aactatgctt tctgaaaaaa
5941 ggcttaggca tctcccatgg caggaagaag cggaagcacc gacgaggaac tcctcagagc
       rev cds start ->
6001 cgtaaggatc atcaatatcc tataccagag cagtaagtaa taagtatatg taatgcacct
                    tat, rev exon end/\intron start
                                                  vpu cds start ->
6061 ttggaaatta gtgcaatagt aggactgtta gtagcgctaa tcttagcaat agtagtgtgg
       /\              <- premature termination of vpu
        1 bp deletion          1 bp deletion, frameshift
6121 actatagtag ctatagaatt taagaaaata ctaaggcaga gaaaaataga caggttagtt
6181 aagagaataa gagaaagaga agaagacagt ggaaatgaga gtgaaggaga cacagataaa
                                env cds ->
                     signal peptide start
6241 ttggccaaac ttgtggaaat gggggacttt gatccttggg ttggtgataa tgtgtagtgc
                                                  env signal peptide/\gp120
6301 ctcaaacaac ttgtgggtta cagtttacta tggggttcct gtgtggagag atgcagatac
6361 caccctattt tgtgcatcag atgccaaagc acatgagaca gaagtgcaca atgtctgggc
6421 cacacatgcc tgtataccca cagaccccaa cccacaagaa atgcacctgg aaaatgtaac
6481 agaaaatttt aacatgtgga aaaataacat ggtagagcag atgcaggagg atgtaatcag
6541 tttatgggat caaagtctaa agccatgtgt aaagttaact cctctctgcg ttactttaaa
6601 ttgtaccaat gctaattgga ccaatgctaa tgtgaccaat gtcaataaca atgtcactaa
6661 catagtagga aatataacag aggaagtaag aaactgctct tttaatatga ccacagaact
6721 aatagataag aagcagaagg tctatgcact tttttataag cttgatataa gacaaatgaa
6781 tagtaatagt agtgagtata ggttaataaa ttgtaatact tcagtcatta agcaggcttg
6841 tccaaaggta tcctttgacc caattcctat acattattgt actccagctg gttatgcgat
6901 tataaagtgt aatgataaga atttcaatgg gacagggcca tgtaaaaatg tcagctcagt
6961 acaatgcaca catggaatta agccagtagt atcaactcaa ttgctgttaa atggcagtct
7021 agcagaagaa aagataataa tcagatctga aaatctcaca aacaatgcca aaaccataat
7081 agtgcacctt catgaatctg tagaaatcaa ttgtaccaga cccttctaca ataaaagaac
                           V3 loop start ->
7141 aaggacatct ataggacaag gacgagtact ctatagaaca ggagacataa caggaaatat
7201 aggaaaacca tattgtgaga ttaatggaac aaaatggaat aaagttttaa atcaggtaac
                       <- V3 loop end
7261 tgaaaaatta aaagagcact ttaataatag gaacataagc tttcagccac catcaggagg
7321 agatctagaa attacaatgc atcattttat ttgtagaggg gaatttttct attgcaatac
7381 aacacgactg tttaataata cttgcatagg aaataaaacc atgaaggagt gtaatgacac
7441 tatcatactt ccatgcaaga taaagcaaat tataaacatg tggcagggag taggacaagc
7501 aatgtataat cctcccatca gtggaaacat taattgtgta tcaaatatta caggaatact
7561 attgacaaga gatggtggtg gtggtaatgg tacgaataat gaggagacct tcagacctgg
7621 aggagggaac atgaaggaca attggagaaa tgaattatat aaatataaag tagtagaaat
7681 tgaaccacta ggaatagcac ccaccaaggc aaagagaaga gtggtggaga gagaaaaaag
7741 agcagtggga ataggagcta tgatctttgg gttcttagga gcagcaggaa gcactatggg
env gp120/\env gp41
7801 cgcggcgtca ataacgctga cggtacaggc cagacaatta ttgtctggta tagtgcaaca
7861 gcaaagcaat ttgctgaggg ctatagaggc gcagcagcat ctgttgcaac tcacagtctg
7921 gggcattaaa cagctccagg caagagtcct ggctgtggaa agatacctaa aagatcaaaa
7981 gttcctagga ctttggggct gctctggaaa aatcatctgc accactgctg tgccctgaaa
8041 ctccacttgg agtaatagat cttatgaaga gatttgggac aacatgacat ggatagaatg
8101 ggagagagaa attagcaatt acacaaacca aatatatgag atacttacag aatcgcagaa
8161 ccagcaggaa agaaatgaaa aggacttgtt agaattggat aaatgggcaa gtctgtggaa
8221 ttggtttggc ataacaaaat ggctgtggta tataaaaata tttataatga tagtaggagg
```

```
      8281 tttaataggt ttaagaataa tttttgctgt gctttctata gtaaatagag ttaggcaggg
      8341 atactcacct ttgtctttcc agacccctac ccatcatcag agggggggaac ccgacagacc
               tat, rev intron end/\exon start
      8401 cgaaagaatc gaagaaggag gtggcgagca aggcagagac agatccgtgc gattagtgag
                                                                    <- tat
cds end
      8461 cggattctta gcacttgcct gggacgatct acggagcctg tgcctcttca gctaccaccg
      8521 cttgagagac ttcatcttga ttgcagcgag gactgtggaa cttctgggac acagcagtct
      8581 caagggactg agacgggggt gggaaggcct caaatatctg gggaatcttc tgttatattg
      8641 gggccaggaa ctaaaaatta gtgctatttc tttgtttgat gctttagcag tagtggtagc
                              <- rev cds end
      8701 ggggtggaca gatagggtta tagaagtagc acaaggagct tggagagcca ttctccacat
      8761 acctagaaga atcagacagg gcttagaaag ggctttgcta taacatggga ggcaagtggt
                                                      <- env cds
                                                         gp41 end
                                              nef cds start ->
      8821 caaaaagtag catagtggga tggcctcagg tcagggaaag aataaagcaa actcctccag
      8881 cagcagaagg agtaggagca gtatctcaag atctagataa acatggagca gtaacaagta
      8941 gtaatatgaa taatgctgat tgtgtctggc tgagagcaca agaggaagag ggggtaggct
      9001 ttccagtcag gccgcaggta cctctaagac caatgactta taagggagct tttgatctta
      9061 gcttcttttt aaaagaaaag gggggactgg aagggctagt ttactccaag aaaagacaag
                 U3 region start 3' LTR ->
      9121 agatccttga cttatgggtc tatcatacac aaggcttctt ccctgattgg cataactaca
      9181 caccagggcc agggatcaga tatccactgt gttttggatg gtgcttcaaa ctagtaccag
      9241 ttgacccaag agaagtagag gaggacaaca aaggagaaaa caactgcctg ttacacccca
      9301 tgagccagca tggaatagag gacgaagaaa gagaagtgct gatatggaag tttgacagtg
      9361 ccctagcacg aagacacata gcccgagaac tgcgtccaga gttctataaa gactgctgac
                                                                    <-
                                                                    nef
cds end
      9421 aaagaagttt ctaactagga cttccgctgg ggactttcca ggggaggtgt ggccggggcg
                                          -- ---------- Nf-Kappa B
      9481 gagttgggga gtggttaacc ctcagatgct gcataaaagc agccgctttt cgcttgtact
      9541 gggtctctct tgttaggcca ggtcgagccc gggagctctc tggctagcag gggaacccac
LTR U3 region/\LTR repeat region
      9601 tgcttaaagc ctcaataaag cttgccttga gtgcttaaag tggtgtgtgc ccatctgtgt
                           ------        LTR R repeat/\LTR U5 region
                        poly-A signal
      9661 taggactctg gtaactagag atccctcaga tcactctaga ctgagtaaaa atctctagca
                                                                    <-
                                                                    3'LTR
                                                                    U5 end
```

```
LOCUS       HIVCM240    9203 bp    DNA             VRL       13-SEP-1996
DEFINITION  Human immunodeficiency virus type 1, strain CM240, complete
            proviral genome.
ACCESSION   U54771
SOURCE      Human immunodeficiency virus type 1 (HIV-1), strain CM240; PBMCs
            extracted from male patient from Thailand; proviral DNA amplified;
            complete genomic sequence determined from three clones of viral genome.
  ORGANISM  Human immunodeficiency virus type 1
            Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;
            Retroviridae; Lentivirinae.
REFERENCE   1  (bases 1 to 9203)
  AUTHORS   Carr,J.K., Salminen,M.O., Koch,C., Gotte,D., Artenstein,A.W.,
            Hegerich,P.A., St. Louis,D., Burke,D.S. and McCutchan,F.E.
  TITLE     Full-length sequence and mosaic structure of a human
            immunodeficiency virus type 1 isolate from Thailand
  JOURNAL   J. Virol. 70 (9), 5935-5943 (1996)
REFERENCE   2  (bases 1 to 9203)
  AUTHORS   Carr,J.K., Salminen,M.O., Gotte,D.R., Artenstein,A.W.,
            Hegerich,P.A., St., Louis,D., Burke,D.S. and McCutchan,F.E.
   TITLE    Direct Submission
 JOURNAL    Submitted (11-APR-1996) J.K. Carr, Molecular Epidemiology, Henry M.
            Jackson, 1600 E. Gude Dr., Rockville, MD 20850, USA
 COMMENT    Blood from an asymptomatic 21-year-old Thai man was transported
            from Chiang Mai to the USA where PBMCs were separated and
            co-cultivated with PHA-stimulated donor PBMCs.  DNA from p24
            antigen-positive culture was used to amplify the proviral DNA.
            The complete genomic sequence of the provirus was determined by the
            compilation of three clones containing different parts of the viral
            genome.  CM240 is an example of a Thai E virus, which is a mosaic of
            a clade A virus and a clade E virus.
FEATURES             Location/Qualifiers
     source          1..9203
                     /organism="Human immunodeficiency virus type 1"
                     /proviral
                     /strain="CM240"
                     /note="Blood from an asymptomatic young Thai man was
                     transported from Chiang Mai to the USA where PBMCs were
                     separated and co-cultivated with PHA-stimulated donor
                     PBMCs.  DNA from p24 antigen-positive culture was used to
                     amplify the proviral DNA.  The complete genomic sequence
                     of the provirus was determined by the compilation of three
                     clones containing different parts of the viral genome.
                     Once the sequence of the provirus was determined the
                     genome was permuted to the organizational structure of the
                     viral genomic RNA; CM240 is an example of a Thai E virus,
                     which is a mosaic of a clade A virus and a clade E virus.
                     See also GenBank accession numbers L14572 and L11761."
     source          join(1..180,8986..9203)
                     /organism="Human immunodeficiency virus type 1"
                     /clone="1"
     source          181..8369
                     /organism="Human immunodeficiency virus type 1"
                     /clone="2"
     CDS             355..1866
                     /gene="gag"
                     /codon_start=1
                     /product="gag"
                     /db_xref="PID:g1537051"
                     /translation="MGARASVLSGGKLDAWEKIRLRPGGRKKYRLKHLVWASRELERF
```

```
                ALNPSFLETAEGCQQIIEQLQSTLKTGLEELKSLFNTVATLWCVHQRIEVKDTKEALD
                KIEEVQNKSQRKTQQAAAGTGSSSKVSQNYPIVQNAQGQMAHQPLSPRTLNAWVKVVE
                EKGFNPEVIPMFSALSEGATPQDLNMMLNIVGGHQAAMQMLKETINEEPAEWDRVHPV
                HAGPIPPGQMREPRGSDIAGTTSTLQEQIGWMTNNPPIPVGDIYKRWIILGLNKIVRM
                YSPVSILDIRQGPKEPFRDYVDRFYKTLRAEQATQEVKNWMTETLLVQNANPDCKSIL
                KALGTGATLEEMMTACQGVGGPSHKARVLAEAMSHAQHATIMMQRGNFKGQKRIKCFN
                CGREGHLARNCRAPRKQGCWKCGKEGHQMKDCTERQANFLGKIWPSNKGRPGNFPQSR
                PEPTAPPAENWGMGEEITGEEITSLPKQEQKDKEHPPPLVSLKSLFGNDPLSQ"
CDS             <1644..4670
                /gene="pol"
                /note="Method: conceptual translation supplied by author"
                /codon_start=1
                /product="pol"
                /db_xref="PID:g1537052"
                /translation="FFRENLAFQQGKAGEFSSEQTRANSPTSRKLGDGGRDNGGRDNL
                LTEAGAERQGTSSSFSFPQITLWQRPLVTVKIGGQLKEALLDTGADDTVLEDINLPGK
                WKPKMIGGIGGFIKVKQYDQILIEICGKKAIGTVLVGPTPVNIIGRNMLTQIGCTLNF
                PISPIDTVPVTLKPGMDGPKVKQWPLTEEKIKALTEICKEMEEEGKISKIGPENPYNT
                PVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYF
                SVPLDESFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRIK
                NPEMVIYQYKDDLYVGSDLEIGQHRTKIEELRAHLLSWGFTTPDKKHQKEPPFLWMGY
                ELHPDRWTVQPIELPEKDSWTVNDIQKLVGKLNWASQIYAGIKVKQLCKLLRGAKALT
                DIVPLTEEAELELAENREILKTPVHGVYYDPSKDLVAEVQKQGQDQWTYQIYQEPFKN
                LKTGKYARRGSAHTNDVRQLTEVVQKVATESIVIWGKTPKFRLPIQRETWETWWMEYW
                QATWIPEWEFVNTPPLVKLWYQLEKDPIVGAETFYVDGAASRETKLGKAGYVTDRGRQ
                KVVSLTETTNQKTELHAIHLALQDSGSEVNIVTDSQYALGIIQAQPDRSESEVVNQII
                EELIKKEKVYLSWVPAHKGIGGNEQVDKLVSSGIRKVLFLDGIDKAQEEHERYHSNWR
                TMASDFNLPPIVAKEIVTNCDKCQLKGEAMHGQVDCSPGIWQLDCTHLEGKVILVAVH
                VASGYIEAEVIPAETGQETAYFLLKLAGRWPVKVIHTDNGSNFTSAAVKAACWWANVQ
                QEFGIPYNPQSQGVVESMNKELKKIIGQVREQAEHLKTAVQMAVFIHNFKRKGGIGGY
                SAGERIIDIIATDIQTKELQKQITKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAVVIQ
                DNSDIKVVPRRKAKIIRDYGKQMAGDDCVAGRQDED"
CDS             4615..5187
                /gene="vif"
                /codon_start=1
                /product="vif"
                /db_xref="PID:g1537053"
                /translation="MENRWQVMIVWQVDRMRIRTWNSLVKHHMYISKKAKKWFYRHHY
                ESQHPKVSSEVHIPLGEARLVIRTYWGLQTGEKDWHLGHGVSIEWRQRKYSTQIDPDL
                ADKLIHLQYFGCFSDSAIRKAILGQVVRRRCEYPSGHNKVGSLQYLALKALTTPKRIR
                PPLPSVEITEDRWNKPQKRGHRENPTMNGH"
CDS             5130..5417
                /gene="vpr"
                /codon_start=1
                /product="vpr"
                /db_xref="PID:g1537054"
                /translation="MEQAPEEGPQREPYNEWTLELLEELKNEAVRHFPRPWLQGLGQY
                IYNNYGDTWEGVEAIIRMLQQLLFVHFRIGCQHSRIGIMPGRRGRNGTGRS"
CDS             join(5398..5612,7932..8022)
                /gene="tat"
                /codon_start=1
                /product="tat"
                /db_xref="PID:g1537055"
                /translation="MELVDPNLEPWNHPGSQPTTACSKCYCKKCCWHCQLCFLKKGLG
                ISYGRKKRKHRRGTPQSSKDHQNPIPKQPLPIIRRNPTDPKESKKEVASKAETDQCD"
CDS             join(5537..5612,7932..8227)
                /gene="rev"
                /codon_start=1
                /product="rev"
```

```
                     /db_xref="PID:g1537056"
                     /translation="MAGRSGSTDEELLRAVRIIKILYQSNPFPSSEGTRQTRRNRRRR
                     WRARQRQISAISERILSTCLGRSTEPVPLQLPPVERLNLDCSEDGGTSGTQQSQGTET
                     GVGRPQISGESSVILGPGTKN"
     CDS             5633..5701
                     /gene="vpu"
                     /note="this clone has a premature translation stop in vpu"
                     /codon_start=1
                     /product="truncated vpu"
                     /db_xref="PID:g1537057"
                     /translation="MTPLEISAIVGLIVALILAIVV"
     CDS             5796..8369
                     /gene="env"
                     /codon_start=1
                     /product="env"
                     /db_xref="PID:g1537058"
                     /translation="MRVKETQMNWPNLWKWGTLILGLVIICSASDNLWVTVYYGVPVW
                     RDADTTLFCASDAKAHETEVHNVWATHACVPTDPNPQEIHLENVTENFNMWKNNMVEQ
                     MQEDVISLWDQSLKPCVKLTPLCVTLNCTNANLTNGSSKTNVSNIIGNITDEVRNCTF
                     NMTTELTDKKQKVHALFYKLDIVQIEDKKTSSEYRLINCNTSVIKQACPKISFDPIPI
                     HYCTPAGYAILKCNDKNFNGTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAEEEIIIR
                     SEDLTNNAKTIIVHLNKSVEINCTRPSNNTRTSITIGPGRVFYRTGDIIGNIRKAYCE
                     INGTKWNKVLKQVTEKLKEHFNKTIIFQPPSGGDLEITMHHFNCRGEFFYCNTTKLFN
                     NTCLGNETMAGCNDTITLPCKIKQIINMWQGAGQAMYAPPISGRINCVSNITGILLTR
                     DGGVNNTDNETFRPGGGNIKDNWRSELYKYKVVQIEPLGIAPTRAKRRVVEREKRAVG
                     IGAMIFGFLGAAGSTMGAASITLTVQARQLLSGIVQQQSNLLRAIEAQQHLLQLTVWG
                     IKQLQARVLAVERYLKDQKFLGLWGCSGKIICTTAVPWNSTWSNRSFEEIWNNMTWIE
                     WEREISNYTNQIYEILTESQNQQDRNEKDLLELDKWASLWNWFDITNWLWYIKIFIMI
                     VGGLIGLRIIFAVLSIVNRVRQGYSPLSFQTPSHHQKEPDRPEGIEEGGGEQGRDRSV
                     RLVSGFLALAWDDLRSLCLFSYHRLRDLTLIAARTVELLGHSSLKGLRRGWEGLKYLG
                     NLLLYWGQELKISAISLLDATAIAAAGWTDRVIEVAQGAWRAILHIPRRIRQGLERTL
                     L"
     source          8370..8985
                     /organism="Human immunodeficiency virus type 1"
                     /clone="3"
     CDS             8371..8985
                     /gene="nef"
                     /codon_start=1
                     /product="nef"
                     /db_xref="PID:g1537059"
                     /translation="MGSKWSKSSIVGWPQVREKIKQTPPATEGVGAVSQDLDKHGAIT
                     SSNIDNADCVWLRAQEDEEVGFPVMPQVPLRPMTYKGAFDLSFFLKEKGGLDGLIYSK
                     KRQEILDLWVYNTQGFFPDWQNYTPGPGIRFPLCFGWCFKLVPVDQREVEEDNKGENN
                     CLLHPMSQHGIEDEEREVLMWKFDSALARKHVAREQHPEYYKDC"
     misc_signal     9014..9024
                     /note="NFkappaB"
     misc_signal     9029..9038
                     /note="sp1.3"
     misc_signal     9040..9049
                     /note="sp1.2"
     misc_signal     9051..9061
                     /note="sp1.1"
     TATA_signal     9080..9084
BASE COUNT     3336 a   1622 c   2220 g   2025 t
ORIGIN
        1 GGgtctctct tgttagacca ggtcgagccc gggagctctc tggctagcaa gggaacccac
R rpt start->
 mRNA start->
```

```
  61 tgcttaaagc ctcaataaag cttgccttga gtgcttaaag tggtgtgtgc ccgtctgtgt
                    ------            R repeat/\LTR U5
               poly-A signal
 121 taggactctg gtaactagag atccctcaga ccactctaga ctgagtaaaa atctctacca
                                                                  <- 5' LTR
 181 gtggcgcccg aacagggcac tcgaaagcga aagttaatag ggactcgaaa gcgaaagttc
 241 cagagaagtt ctctcgacgc aggactcggc ttgctgaggt gcacacagca agaggcgaga
 301 gcggcgactg gtgagtacgc caaattttga ctagcggagg ctagaaggag agagatgggt
             /\major 5' sj                          gag cds ->
                                          p17 MA (matrix) start
 361 gcgagagcgt cagtattaag tgggggaaaa ttagatgcat gggaaaaaat tcggttgcgg
 421 ccaggggaa gaaaaaaata taggctgaaa catttagtat gggcaagcag agagttagaa
 481 agattcgcac ttaaccctag ctttttagaa acagcagaag gatgtcaaca aataatagaa
 541 cagttacagt caactctcaa gacaggatta gaagaactta aatcattatt taatacagta
 601 gcaaccctct ggtgcgtaca ccaaaggata gaggtaaaag acaccaagga agctttagat
 661 aaaatagagg aagtacaaaa taagagccag cgaaagacac agcaggcagc agctggcaca
 721 ggaagcagca gcaaagtcag ccaaaattac cctatagtgc aaaatgcaca agggcaaatg
                           gag p17 MA/ \gag p24 CA
 781 gcacatcagc ctttatcacc tagaactttg aatgcatggg tgaaagtagt agaagaaaag
 841 ggttttaacc cagaagtaat acccatgttc tcagcattat cagagggagc caccccacaa
 901 gatttaaata tgatgctaaa tatagtgggg ggacaccagg cagcaatgca aatgttaaaa
 961 gaaaccatca atgaggaacc tgcagaatgg gatagggtac acccagtaca tgcagggcct
1021 attccaccag gccagatgag ggaaccaagg ggaagtgaca tagcaggaac tactagtacc
1081 cttcaagaac aaataggatg gatgacaaac aatccaccca tcccagtggg agacatctat
1141 aaaaggtgga taatcctggg attaaataaa atagtaagaa tgtatagccc tgttagcatt
1201 ttggacataa gacaagggcc aaaagaaccc ttcagagact atgtagatag gttctataaa
1261 actctcagag cggaacaagc tacacaggag gtaaaaaact ggatgacaga aaccttgcta
1321 gtccaaaacg cgaatccaga ctgtaagtcc attttaaaag cattaggaac aggagctaca
1381 ttagaagaaa tgatgacagc atgccaggga gtgggaggac ctagccataa agcaagggtt
1441 ttggctgaag caatgagcca cgcacaacat gcaactataa tgatgcagag aggcaatttc
gag p24 CA/\gag p2x                          gag p2x/\gag p7NC
                                                    \gag-pol TF
1501 aagggccaga aaagaattaa gtgcttcaac tgtggtagag aaggacacct agccagaaat
1561 tgcagggccc ctagaaaaca gggttgttgg aaatgcggga aggaaggaca tcaaatgaaa
1621 gactgcactg agagacaggc taatttttta gggaaaattt ggccttccaa caagggaagg
                              ------              ^^^^^^^^       ^^^^^^
                 gag-pol  ribosomal slip site      stem   loop    stem
1681 ccggggaatt ttcctcagag cagaccagag ccaacagccc caccagcaga aaactggggg
     ^^  gag p7NC/\gag p6
1741 atgggggaag agataacggg ggaagagata acctccttac cgaagcagga gcagaaagac
1801 aaggaacatc ctcctccttt agtttccctc aaatcactct ttggcaacga ccccttgtca
               gag-pol fusion TF/\pol protease
1861 cagtaaaaat aggaggacag ctgaaagaag ctctattaga tacaggagca gatgatacag
        <- gag cds
           p6 end
1921 tattagaaga tataaatttg ccaggaaaat ggaaaccaaa aatgataggg ggaattggag
1981 gttttatcaa ggtaaagcaa tatgatcaga tacttataga aatctgtgga aaaaaggcta
2041 taggtacagt attagtagga cctacacctg tcaacataat tggacgaaat atgttgactc
2101 agattggttg tactttaaat ttcccaatta gtcctattga cactgtacca gtaacattaa
                pol protease/\pol reverse transcriptase
2161 agccaggaat ggatggacca aaggttaaac agtggccatt gacagaagaa aaaataaaag
2221 cattaacaga aatttgtaaa gagatggaag aggaaggaaa aatctcaaaa attgggcctg
2281 aaaatccata caatactcca gtatttgcta taaagaaaaa ggacagcacc aaatggagga
2341 aattagtaga tttcagagag ctcaataaaa gaactcagga cttttgggaa gttcaattag
2401 gaataccgca tccagcaggt ttaaaaaaga aaaaatcagt aacagtacta gatgtgggag
2461 atgcatattt ttcagttcct ttagatgaaa gctttagaaa gtatactgca ttcaccatac
2521 ctagtataaa caatgagaca ccaggaatca gatatcagta caatgtgctg ccacagggat
2581 ggaaaggatc accggcaata ttccagagta gcatgacaaa aatcttagag ccctttagaa
2641 taaaaaatcc agaaatggtt atctatcaat acaaggatga cttgtatgta ggatctgatt
```

```
2701 tagaaatagg gcagcacaga acaaaaatag aggagctaag agctcatcta ttgagctggg
2761 gatttactac accagacaaa aagcatcaga aggaacctcc attcctttgg atgggatatg
2821 aactccatcc tgacagatgg acagtccagc ctatagaact gccagaaaaa gacagctgga
2881 ctgtcaatga tatacagaaa ttagtgggaa aactaaattg ggcaagtcaa atttatgcag
2941 ggattaaggt aaagcaactg tgtaaactcc tcaggggagc taaagcacta acagacatag
3001 taccactgac tgaagaagca gaattagagt tggcagagaa cagggagatt ctaaaaaccc
3061 ctgtgcatgg agtatattat gacccatcaa aagacttagt agcagaagta cagaaacaag
3121 ggcaggacca atggacatat caaatttatc aagagccatt taaaaatcta aaaacaggaa
3181 aatatgccag aagagggtct gctcacacta atgatgtaag acaattaaca gaagtggtgc
3241 aaaaagtagc cacagaaagc atagtaatat ggggaaagac ccctaaattt agactaccca
3301 tacaaagaga aacatgggaa acatggtgga tggagtattg gcaggctacc tggattcctg
3361 aatgggagtt tgttaatacc cctcctctag taaaattatg gtaccaatta gaaaaagacc
3421 ccatagtagg agcagagact ttctatgtag atggggcagc tagtagggag actaagctag
    pol reverse transcriptase/\pol RNAse
3481 gaaaagcagg gtatgtcact gacagaggaa gacaaaaggt agtttcccta actgagacaa
3541 caaatcaaaa gactgaatta catgcgatcc atttagcctt gcaggattca ggatcagaag
3601 taaatatagt aacagactca caatatgcat taggaatcat tcaggcacaa ccagacagga
3661 gtgaatcaga agtagtcaac caaataatag aggagctaat aaaaaaggag aaagtctacc
3721 tgtcatgggt accagcacac aaggggattg gaggaaatga acaagtagat aaattagtca
3781 gttcaggaat caggaaggtg ctatttttag atgggataga taaggctcaa gaagaacatg
                    pol RNAse/\pol integrase
3841 aaagatatca cagcaattgg agaacaatgg ctagtgattt taatttgcca cctatagtag
3901 caaaggaaat agtaaccaac tgtgataaat gtcaactaaa aggggaagct atgcatggac
3961 aagtagactg tagtccaggg atatggcaat tagattgcac acatctagaa ggaaaagtca
4021 tcctggtagc agtccacgtg gccagtggat atatagaagc agaagttatc ccagcagaaa
4081 caggacagga gacagcatac tttctgctaa aactagcagg aagatggcca gtaaaagtaa
4141 tacacacaga caacggtagc aatttcacca gcgctgcagt taaagcagcc tgttggtggg
4201 ccaatgtcca acaggaattt gggatcccct acaatcccca aagtcaagga gtagtagaat
4261 ctatgaataa ggaattaaag aaaatcatag ggcaggtaag agagcaagct gaacacctta
4321 aaacagcagt acaaatggca gtattcattc acaattttaa aagaaaaggg gggattgggg
4381 ggtacagtgc aggggaaaga ataatagaca taatagcaac agacatacaa actaaagaat
4441 tacaaaaaca aattacaaaa attcaaaatt ttcgggttta ttacagggac agcagagacc
                                                    /\ 3' sj
4501 caatttggaa aggaccagca aaactactct ggaaaggtga aggggcagta gtaatacaag
                                          /\ 5' sj
4561 acaatagtga tataaaagta gtaccaagaa gaaaagcaaa gatcattagg gattatggaa
                                           vif cds start ->
4621 aacagatggc aggtgatgat tgtgtggcag gtagacagga tgaggattag aacatggaac
                                                   <- pol cds
                                                      integrase end
4681 agtttagtaa aacatcatat gtatatctca aagaaagcta aaaagtggtt ttatagacat
4741 cattatgaaa gccagcatcc aaaggtaagc tcagaagtac atatcccact aggagaggct
4801 agattagtaa taagaacata ttggggtctg caaacaggag aaaaggactg gcacttgggt
4861 catggagtct ccatagaatg gaggcagaga aaatatagca cacaaataga tcctgaccta
4921 gcagacaaac tgattcatct acaatatttt ggctgttttt cagactctgc cataaggaaa
                                                  /\ 3' sj
4981 gccatattag gacaagtagt tagacgtagg tgtgaatatc catcaggaca taacaaggta
                                                                /\ 5'sj
5041 ggatccctac aatatttggc actgaaagca ttaacaacac caaaaaggat aaggccacct
5101 ctgcctagtg tagaaataac agaagataga tggaacaagc cccagaagag gggccacaga
                     vpr cds start - >
5161 gagaacccta caatgaatgg acattagaac tgttagagga gcttaaaaat gaagctgtta
                                 <- vif cds end
5221 gacattttcc taggccctgg ctccaaggct taggacagta catctataac aattatgggg
5281 atacttggga aggggttgaa gctataataa gaatgttgca acaactactg tttgttcatt
5341 tcagaattgg gtgtcaacat agcagaatag gcattatgcc agggagaaga ggcaggaatg
       /\ 3' sj                                    tat cds start ->
5401 gaactggtag atcctaacct agagccctgg aatcatccgg gaagtcagcc tacaactgct
                    <- vpr cds end
```

```
5461 tgtagcaagt gttactgtaa aaaatgttgc tggcattgcc aactatgctt tctgaaaaaa
5521 ggcttaggca tctcctatgg caggaagaag cggaagcacc gacgaggaac tcctcagagc
       rev cds start ->     /\ 3' sj
5581 agtaaggatc atcaaaatcc tataccaaag cagtaagtaa taagtatatg taatgacacc
                  (tat,rev,nef) 5' sj /\       vpu cds start ->
5641 tttggaaatt agtgcaatag taggactgat agtagcgcta atcttagcaa tagtagtgtg
5701 aactatagta gctatagaag ttaagaaaat actaaggcaa agaaaaatag acaggttagt
    <- vpu cds premature end
5761 taagagaata agagaaagag cagaagacag tggaaatgag agtgaaggag acacagatga
                                 env cds ->
                        signal peptide start
5821 attggccaaa cttgtggaaa tgggggactt tgatccttgg gttggtgata atttgtagtg
                                                  env signal peptide/\gp120
                                                                <- vpu cds
                                                                   normal end
5881 cctcagacaa cttgtgggtt acagtttatt atggggtgcc tgtgtggaga gatgcagata
5941 ccaccctatt ttgtgcatca gatgccaagg cacatgagac agaagtgcac aatgtctggg
6001 ccacacatgc ctgtgtaccc acagacccca acccacaaga aatacacctg gaaaatgtaa
6061 cagaaaattt taacatgtgg aaaaataaca tggtagagca gatgcaggag gatgtaatca
6121 gtttatggga tcaaagtcta aagccatgtg taaagttaac tcctctctgc gttactttaa
6181 attgtaccaa tgctaatttg accaatggca gtagcaaaac caatgtctct aacataatag
6241 gaaatataac agatgaagta agaaactgta cttttaatat gaccacagaa ctaacagata
6301 agaagcagaa ggtccatgca ctcttttata agcttgatat agtacaaatt gaagataaga
6361 agactagtag tgagtatagg ttaataaatt gtaatacttc agtcattaag caggcttgtc
6421 caaagatatc ctttgatcca attcctatac attattgtac tccagctggt tatgcgattt
6481 taaagtgtaa tgataagaat ttcaatggga cagggccatg taaaaatgtc agctcagtac
6541 aatgcacaca tggaattaag ccagtggtat caactcaatt gctgttaaat ggcagtctag
6601 cagaagaaga gataataatc agatctgaag atctcacaaa caatgccaaa accataatag
6661 tgcaccttaa taaatctgta gaaatcaatt gtaccagacc ctccaacaat acaagaacaa
                       V3 loop start - >
6721 gtataactat aggaccagga cgagtattct atagaacagg agatataata ggaaatataa
6781 gaaaagcata ttgtgagatt aatggaacaa aatggaataa agttttaaaa caggtaactg
                  <- V3 loop end
6841 aaaaattaaa agagcacttt aataagacaa taatctttca accaccctca ggaggagatc
6901 tagaaattac aatgcatcat tttaattgta gaggggaatt tttctattgc aatacaacaa
6961 aactgtttaa taatacttgc ctaggaaatg aaaccatggc ggggtgtaat gacactatca
7021 cacttccatg caagataaag caaattataa acatgtggca gggagcagga caagcaatgt
7081 atgctcctcc catcagtgga agaattaatt gtgtatcaaa tattacagga atactattga
7141 caagagatgg tggtgttaat aatacggata acgagacctt cagacctgga ggaggaaaca
7201 taaaggacaa ttggagaagt gaattatata aatataaagt agtacaaatt gaaccactag
7261 gaatagcacc caccagggca aagagaagag tggtggagag agaaaaaagg gcagtgggaa
                                                 env gp120/ \env gp41
7321 taggagctat gatctttggg ttcttaggag cagcaggaag cactatgggc gcggcgtcaa
7381 taacgctgac ggtacaggcc agacaattat tgtctggtat agtgcaacag caaagcaatt
7441 tgctgagggc tatagaggcg cagcagcatc tgttgcaact cacagtctgg ggcattaaac
7501 agctccaggc aagagtcctg gctgtggaaa gatacctaaa ggatcaaaag ttcctaggac
7561 tttggggctg ctctggaaaa atcatctgca ccactgctgt gccctggaac tccacttgga
7621 gtaatagatc ttttgaagag atttggaaca acatgacatg gatagaatgg gagagagaaa
7681 ttagcaatta cacaaaccaa atatatgaga tacttacaga atcgcagaac caacaggaca
7741 ggaatgaaaa ggatttgtta gaattggata aatgggcaag tctgtggaat tggtttgaca
7801 taacaaattg gctgtggtat ataaaaatat ttataatgat agtaggaggt ttaataggtt
7861 taagaataat ttttgctgtg ctttctatag taaatagagt taggcaggga tactcacctt
7921 tgtctttcca gaccccttcc catcatcaga aggaacccga cagacccgaa ggaatcgaag
               /\ (tat, rev, nef) 3' sj
7981 aaggaggtgg cgagcaaggc agagacagat cagtgcgatt agtgagcgga ttcttagcac
                                                    <- tat cds end
8041 ttgcctggga cgatctacgg agcctgtgcc tcttcagcta ccaccggttg agagacttaa
8101 ccttgattgc agcgaggacg gtggaacttc tgggacacag cagtctcaag ggactgagac
8161 gggggtggga aggcctcaaa tatctgggga atcttctgtt atattggggc caggaactaa
```

```
8221 aaattagtgc tatttctttg cttgatgcta cagcaatagc agcagcgggg tggacagaca
         <-rev cds end
8281 gggttataga agtagcacaa ggagcttgga gagccattct ccacatacct agaagaatca
8341 gacagggctt agaaaggact ttgctataac atgggaagta agtggtcaaa aagtagcata
                                  <- env cds
                                     gp41 end
                          nef cds start ->
8401 gtgggatggc ctcaggtcag ggaaaaaata aagcaaactc ctccagcaac agaaggagta
8461 ggagcagtat ctcaagatct agataaacat ggagcaataa caagtagtaa tatagataat
8521 gctgattgtg tctggctgag agcacaagag gacgaggagg taggctttcc agtcatgccg
8581 caggtacctc taagaccaat gacttataag ggagcttttg atcttagctt ctttttaaaa
8641 gaaaaggggg gactggatgg gctaatttac tccaagaaaa gacaagagat ccttgactta
                        \3'LTR U3 start
8701 tgggtctata atacacaagg cttcttccct gattggcaaa actacacacc agggccaggg
8761 atcagattcc cactgtgttt tggatggtgc ttcaagctag taccagttga ccaaagagaa
8821 gtagaggagg acaacaaagg agaaaacaac tgcctgttac accccatgag ccagcatgga
8881 atagaggacg aagaaagaga agtgctgatg tggaagtttg acagtgccct agcacgaaaa
8941 cacgtagccc gagaacagca tccagagtac tataaagact gctgacaagg aagtttctac
                                                    <- nef cds end
9001 tagaacttcc gctggggact ttccagggga ggtgtggccg gggcggagtt ggggagtagc
                    ------- ---- Nf-Kappa B
9061 taaccctcag atgctgcata aaagcagccg cttttcgctt gtactgggtc tctcttgtta
                                                 LTR U3/\R rpt
9121 gaccaggtcg agcccgggag ctctctggct agcaagggaa cccactgctt aaagcctcaa
                                                                      --
9181 taaagcttgc cttgagtgct taa
     ----                    <- mRNA end
 signal poly-A      R rpt end/
```

```
LOCUS       HIVETH2220   9031 bp    DNA             VRL       19-JUN-1996
DEFINITION  HIV-1 isolate C2220 from Ethiopia, subtype C, complete genome.
ACCESSION   U46016
SOURCE      Human immunodeficiency virus type 1 (HIV-1), strain C2220 (ETH2220),
            clone pNOTA/C2220/C3.6.4; subtype C from Ethiopia; isolate by
            short-term co-culture on patient's PBMCs.
  ORGANISM  Human immunodeficiency virus type 1
            Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;
            Retroviridae; Lentivirinae.
REFERENCE   1  (bases 1 to 9031)
  AUTHORS   Salminen,M.O., Johansson,B., Sonnerborg,A., Ayehunie,S., Gotte,D.,
            Leinikki,P., Burke,D.S. and McCutchan,F.E.
  TITLE     Full-length sequence of an Ethiopian human immunodeficiency virus
            type 1 (HIV-1) isolate of genetic subtype C
  JOURNAL   AIDS Research and Human Retroviruses 12 (14), 1329-1339 (1996)
REFERENCE   2  (bases 1 to 9031)
  AUTHORS   Salminen,M.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1996) Mika O. Salminen, Global Molecular
            Epidemiology, Henry M. Jackson Foundation, 1600 East Gude Drive,
            Rockville, MD 20850, USA
 COMMENT    U46016 is the first reported almost full length subtype C
            sequence from Ethiopia.  In its genomic organization, this clone
            closely resembles subtype A, B, and D isolates except that the core
            promoter contains three potential binding sites for the
            transcription factor NF-kB instead of containing two.
            See also GenBank accession number U15061.
FEATURES             Location/Qualifiers
     source          1..9031
                     /organism="Human immunodeficiency virus type 1"
                     /strain="C2220 (ETH2220)"
                     /note="subtype C from Ethiopia; isolated by short term
                     co-culture on normal human PBMC"
                     /clone="pNOTA/C2220/C3.6.4"
     CDS             170..1684
                     /gene="gag"
                     /note="GAG p17/p24/NCp7/p6"
                     /codon_start=1
                     /product="matrix, capsid and core proteins"
                     /db_xref="PID:g1353861"
                     /translation="MGARASILRGEKLDAWEKIKLRPGGKKHYMLKHLVWANRELEKF
                     ALNPDLLDTSAGCKQIIKQLQPALQTGTEELKSLFNTVATLYCVHQKIEIKDTKEALD
                     KIEEEQNESQQKTQQAGAADRGKDSQNYPIVQNMQGQMVHQPISARTLNAWVKVVEEK
                     AFSPEVIPMFTALSEGATPQDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDRLHPVHA
                     GPVAPGQMRDPRGSDIAGTTSTLQEQIAWMTGNPPVPVGDIYKRWIILGLNKIVRMYS
                     PVSILDIKQGPKEPFRDYVDRFFKTLRAEQATQDVKNWMTDTLLVQNANPDCKTILRA
                     LGPGASLEEMMTACQGVGGPAHKARVLAEAMSQVNNTTIMMQKSNFKGPKRAIKCFNC
                     GKEGHLARNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGRLWPSNKGRPGNFLQSRP
                     EPTAPPESLRPEPTAPPPESFRFEEATPSPKQELKDREALTSLKSLFGNDHLLQ"
     CDS             <1456..4488
                     /gene="pol"
                     /note="initiation codon is unknown"
                     /codon_start=1
                     /product="protease/RT/RNAseH/integrase"
                     /db_xref="PID:g1353862"
                     /translation="FFRETLAFQQGKAREFPSEQTRANSPTRESQTRANSPTTRELQV
                     RGSNTFSEAGAERQGSLNFPQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEINLP
                     GKWKPKMIGGIGGFIKVRQYDQIIIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGRTL
                     NFPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALTAICEEMEQEGKISRIGPENPY
```

```
                NTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDA
                YFSVPLDEGFRKYTAFTIPSTNNETPGIRYQYNVLPQGWKGSPPIFQSSMPQILEPFR
                APNPEIVIYQYMDDLYVGSDLEIGQHRAPIEELREHLLKWGFTTPDKKHQKEPPFLWM
                GYELHPDKWTVQPIQLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGAKA
                LTDIVTLTEEAELELAENREILKEPVHGVFYDPSKDLIAEIQKQGNDQWTFQFYQEPF
                KNLKTGKFAKRGTAHTNDVKQLTAVVQKIALESIVIWGKTPKFRLPIQKETWEAWWTD
                YWQATWIPEWEFVNTPPLVKLWYQLEKEPIAGVETFYVDGAANRETKIGKAGYVTDRG
                RQKIVSLTETTNQKTELQAIQLALQDSGSEVNIVTDSQYALGIILAQPDKSESEIVNQ
                IIEQLISKERVYLSWVPAHKGIGGNEQVDKLVSSGIRKVLFLDGIDKAQEEHEKYHSN
                WRAMANEFNIPPVVPKEIVACCDKCQLKGEAIHGQVNCSPGIWQLDCTHLEGKIILVA
                VHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVRVIHTDNGSNFTSNAVKAACWWAG
                IQQEFGIPYNPQSQGVVESMNKELKKIIGQVREQAEHLKTAVQMAVFIHNFKRRGGIG
                GYSAGERIIDIIASDIQTKELQNQILKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAVV
                IQDNSDIKVVPRRKAKIIRDYGKQMAGADCVAGRQDED"
CDS             4433..5011
                /gene="vif"
                /note="infectivity factor"
                /codon_start=1
                /product="VIF"
                /db_xref="PID:g1353863"
                /translation="MENRWQVLIVWQVDRMKIRTWNSLVKHHMHISRRANGWVYRHHY
                DSRHPKVSSEVHIPLGEARLIIKTYWGLQTGERDWHLGHGVSIEWRLRSYNTQVDPGL
                ADHLIHMHYFDCFAESAIRKAILGYRVSPRCDYQAGHNKVGSLQYLALTALIKPKKAK
                PPLPSVSKLVEDKWNKPQKTRGRRGNHTMNGH"
CDS             4951..5241
                /gene="vpr"
                /codon_start=1
                /product="VPR"
                /db_xref="PID:g1353864"
                /translation="MEQAPEDQGPQREPYNEWALELLEELKQEAVRHFPRPWLHNLGQ
                YIYETYGDTWSGVEALIRTLQQLMFIHFRIGCQHSRIGILRQRRARNGASRS"
CDS             5222..5350
                /gene="tat"
                /note="transactivating factor; prematurely truncated at nt
                5350 due to a deletion at nt 5344; second exon not
                expressed"
                /codon_start=1
                /product="TAT"
                /db_xref="PID:g1353865"
                /translation="MEPVDPNLEPWNHPGSQPKTACNQCYCKKCSYHCLVCFLTKA"
CDS             join(5360..5435,7755..8002)
                /gene="rev"
                /note="regulator of protein expression; prematurely
                truncated at nt 8002"
                /codon_start=1
                /product="REV"
                /db_xref="PID:g1353866"
                /translation="MAGRSGDSDEELLKAVRIIKILYQSNPYPTPEGTRQARRNRRRR
                WRARQRQIHTLSERILSNFLGRPAEPVPLQLPPLERLNLDCSEDSGTSGTQQSQGTTE
                GVGNP"
CDS             5459..5719
                /gene="vpu"
                /codon_start=1
                /function="promotes extracellular release of virions"
                /product="VPU"
                /db_xref="PID:g1353867"
                /translation="MVDLLAKVDYRIVIVAFIVALIIAIVVWTIAYIEYRKLLRQRRI
                DRLIKRTRERAEDSGNESDGDTEELSTMVDMGNLRLLDVNDL"
```

```
     CDS             5637..8192
                     /gene="env"
                     /note="viral envelope gene;  contains external and
                     transmembrane glycoproteins; cleaved to gp120 & gp41"
                     /codon_start=1
                     /product="gp160"
                     /db_xref="PID:g1353868"
                     /translation="MKVMGIQRNCQQWWIWGILGFWMLMICNGMGNLWVTVYYGVPVW
                     KDASPTLFCASDAKAYDTEVHNVWGTFACVPTDPSPQELGLENVTENFNMWKNDMVEQ
                     MHQDIISLWDQGLKPCVKLTPLCVTLNCNAIKNNTKVTNNSINSANDEMKNCSFNITT
                     ELRDKKRKAYALFYKLDIVPLNNGSTDYRLINCNTSTITQACPKVSLDPIPIHYCAPA
                     GYAILKCRDKTFTGTGPCHNVSTVQCTHGIKPVVSTQLLLNGSIAEGETIIRFENLTN
                     NAKIIIVQLNESVEITCTRPSNNTRESIRIGPGQTFYATGDIIGDIRQAHCNISEEKW
                     NKTLQKVKEKLQKHFPNKTIEFKPSSGGDLEITTHSFNCGGEFFYCNTSNLFNSTKLE
                     LFNSSTNLNITLQCRIKQIINMWQGVGRAMYAPPIEGIIMCRSNITGLLLTRDGAKEP
                     HSTKEIFRPEGGDMRDNWRSELYKYKVVEIKPLGVAPTKPKRRVVEREKRAALGALFL
                     GFLGAAGSTMGAASITLTVQARQLLSGIVQQQSNLLKAIEAQQHMLQLTVWGIKQLQT
                     RVLAIERHLRDQQLLGIWGCSGKLICTTAVPWNSSWSNKSQEEIWDNMTWMQWDREIS
                     NYTDIIYNLLEVSQNQQDKNEKDLLALDKWENLWNWFNITNWLWYIKIFIMIVGGVIG
                     LRIIFAVLSIVNRVRQGYSPLSFQTLIPHPRGPDRLGGIEEEGGEQGRDRSIRLVNGF
                     LAIFWDDLRSLCLFSYHRLRDLILIAARTVELLGRSSLKGLQRGWETLKYLGSLVQYW
                     GLELKKSAINLLNTTAIVVGEGTDRFIELIQRIWRAFCNIPRRIRQGLEAALQ"
     misc_RNA        7145..7378
                     /gene="env"
                     /note="RRE region of mRNA; viral envelope gene"
     CDS             8194..8817
                     /gene="nef"
                     /note="regulatory protein"
                     /codon_start=1
                     /product="NEF"
                     /db_xref="PID:g1353869"
                     /translation="MGGTMSKCSPVGWPAIRERIRRAAPAAEGVGAASRDLDKYGALT
                     SSNTPANNPDCAWLEAQEEEEEVGFPVRPQVPLRPMTYKAAFDLSLFLKEKGGLEGLI
                     YSKKRQEILDLWVYNTQGFFPDWQNYTPGPGVRYPLTFGWCFKLVPVDPSEVEEINEG
                     ENNCLLHPASLHGMEDEDREVLKWKFDSHLARRHMARELHPEYYKDC"
     LTR             8486..>9031
                     /note="3', U3 and R regions only"
     protein_bind    8825..8834
                     /note="extra NFkB site III"
                     /bound_moiety="NFkb"
     protein_bind    8838..8847
                     /note="NFkB site II"
                     /bound_moiety="NFkb"
     protein_bind    8851..8860
                     /note="NFkB site I"
                     /bound_moiety="NFkb"
     misc_RNA        8939..8997
                     /note="TAR region of mRNA"
BASE COUNT     3270 a   1616 c   2141 g   2004 t
ORIGIN
        1 aaatctctag cagtggcgcc cgaacagggg acctgaaagc gaaagtgaga ccagaggaga
                     <- 5' LTR
       61 tctctcgacg caggactcgg cttgctgaag tgcactcggc aagaggcgag agcggcgact
      121 ggtgagtacg ccaattttta tttgactagc ggaggctaga aggagagaga tgggtgcgag
          /\ major 5' sj                                  gag cds - >
                                      p17 MA (matrix) start
      181 agcgtcaata ttaagaggcg aaaaattaga tgcctgggaa aaaattaagt taaggccagg
      241 gggaaagaaa cactatatgc tgaaacacct agtctgggca aacagggagc tggaaaaatt
      301 tgcacttaac cctgaccttt tagatacatc agcaggctgt aaacaaataa ttaaacagct
```

```
 361 acaaccagct cttcagacag gaacagagga acttaaatca ttatttaata cagtggcaac
 421 tctctattgt gtacatcaaa agatagagat aaaagacacc aaggaagcct tagacaagat
 481 agaggaagaa caaaacgaaa gtcagcaaaa aacacagcag gcaggagcag ctgacagagg
 541 aaaggacagt caaaattatc ctatagtgca gaatatgcag gggcaaatgg tacatcagcc
              gag p17 MA/\gag p24 CA
 601 catatcagct agaactttga atgcatgggt aaaagtagta gaggaaaagg ctttcagccc
 661 agaggtaata cccatgttta cagctttatc agaaggagcc accccacaag atttaaacac
 721 catgctaaat acagtggggg gacatcaagc agccatgcaa atgttaaaag acaccatcaa
 781 tgaggaggct gcagaatggg acaggttaca tccagtgcat gcagggcctg ttgcaccagg
 841 ccaaatgaga gacccaaggg gaagtgacat agcaggaaca actagtaccc ttcaggaaca
 901 aatagcatgg atgacaggga acccacctgt tccagtggga gacatctata aaagatggat
 961 aatcctgggg ctaaataaaa tagtaagaat gtatagccct gtcagcattt tggacataaa
1021 acaaggacca aaggaacctt ttagagacta tgtagaccgg ttctttaaaa ccttaagagc
1081 tgaacaagct acacaagatg taaaaaattg gatgacagac accttgttgg tccaaaatgc
1141 gaacccagat tgtaaaacca ttttaagagc attagggcca ggggcttcat tagaagaaat
1201 gatgacagca tgtcagggag tgggaggacc tgcccacaaa gcaagagtgt tggctgaggc
                                                  gag p24 CA/\gag p2x
1261 aatgagccaa gtaaacaata caaccataat gatgcagaaa agcaatttta agggccctaa
                               gag p2x/\gag p7NC
                                    \gag-pol TF
1321 aagagcaatt aaatgtttca actgtggcaa ggaagggcac ctagccagaa attgcagggc
1381 ccctaggaaa aaaggctgtt ggaaatgtgg aaaggaagga caccaaatga aagactgtac
1441 cgagagacag gctaattttt tagggagact ttggccttcc aacaagggaa ggccagggaa
                     ----- - gag-pol ribosomal slip site
1501 tttccttcag agcagaccag agccaacagc cccaccagag agtctcagac cagagccaac
gag p7NC/\gag p6
1561 agccccacca ccagagagct tcaggttcga ggaagcaaca ccttctccga agcaggagct
1621 gaaagacagg gaagccttaa cttccctcaa atcactcttt ggcaacgacc acttgttaca
            gag-pol fusion TF/\pol protease
1681 ataaaaatag ggggacagct aaaggaggct ctcttagaca caggagcaga tgatacagta
      <- gag cds
         p6 end
1741 ttagaagaaa taaatttgcc aggaaaatgg aaaccaaaaa tgataggagg aattggaggt
1801 tttattaaag taagacagta tgatcaaata atcatagaaa tttgtggaaa aaaggctata
1861 ggtacagtac tagtaggacc tacacctgtc aacataattg gcagaaacat gttgactcag
1921 cttggacgca cattaaactt tccaattagt cccattgaaa ctgtaccagt aaaattaaag
              pol protease/\pol reverse transcriptase
1981 ccaggaatgg atggcccaaa agtcaaacaa tggccattga cagaagaaaa gataaaagca
2041 ttaacagcaa tttgtgaaga aatggagcag gaaggaaaaa tttcaagaat tgggcctgaa
2101 aacccatata atactccagt atttgccata aaaaagaagg acagtactaa gtggagaaaa
2161 ttagtagatt tcagggaact caataaaaga actcaagact tttgggaagt tcaattaggg
2221 ataccacacc cagcagggtt aaaaaagaaa aaatcagtga cagtactaga tgtgggggat
2281 gcatatttct cagttccttt agatgaaggt ttcagaaaat atactgcatt caccatacct
2341 agtacaaaca atgaaacacc agggattaga tatcaatata atgtcctccc acagggatgg
2401 aaaggatcac caccaatatt ccagagtagc atgccccaaa tcttagagcc ctttagggcc
2461 cccaacccag aaatagttat ctatcaatat atggatgact tgtatgtagg atctgactta
2521 gaaatagggc aacatagagc cccaatagaa gagttaagag aacatctatt aaagtgggga
2581 tttaccacac cagacaagaa acatcagaaa gaacctccat ttctttggat ggggtatgaa
2641 cttcatcctg acaaatggac agtacagcct atacagctgc cagaaaagga tagctggact
2701 gtcaatgata tacaaaagtt agtgggaaaa ttaaactggg caagtcagat ttacccaggg
2761 attaaagtaa ggcaactgtg taaactcctt aggggagcca aagcactaac agacatagta
2821 acactaactg aagaagcaga attagaattg gcagagaaca gggaaattct aaaggaacca
2881 gtacatggag tattttatga cccatcaaaa gacttaatag ctgaaataca gaaacagggg
2941 aatgaccaat ggacatttca attttaccaa gaaccattca aaaatctgaa aacagggaag
3001 tttgcaaaaa gagggactgc ccacactaat gatgtaaagc agttaacagc ggtagtgcaa
3061 aagatagccc tggaaagcat agtaatatgg ggaaagactc ctaaatttag attacccatc
3121 cagaaagaaa catgggaagc atggtggaca gactattggc aagccacctg gattcctgaa
3181 tgggagtttg ttaatacccc tcccctagta aaattatggt accagctaga gaaagaaccc
```

```
3241 atagcaggag tagaaacttt ctatgtagat ggagcagcta atagggaaac taaaatagga
  pol reverse transcriptase/\pol RNAse
3301 aaagcagggt atgttactga tagaggaagg cagaaaattg tttctctaac tgaaacaaca
3361 aatcagaaga ctgaattaca agcgatccag ctagcgttac aggattcagg atcagaagta
3421 aacatagtaa cagactcaca gtatgcatta gggatcatcc tggcacaacc agataagagt
3481 gaatcagaga tagtcaatca aataatagaa cagttaataa gcaaggaaag ggtctacctg
3541 tcatgggtac cagcacataa aggaattgga ggaaatgaac aagtagataa attagtaagt
3601 agtggaatca ggaaagtgct gtttctagat ggaatagata aggctcaaga agagcatgaa
                pol RNAse/\pol integrase
3661 aaatatcaca gcaattggag agcaatggct aatgaattta atatcccacc cgtagtaccc
3721 aaagaaatag tagcttgctg tgataaatgt cagctaaaag gggaagccat acatggacaa
3781 gtaaattgta gtccagggat atggcaatta gattgtacac acttagaagg gaaaatcatc
3841 ctggtagcag tccatgtagc cagtggctac atagaggcag aggttattcc agcagaaaca
3901 ggacaagaaa cagcatactt tctactaaaa ttagcaggga gatggccagt cagggtaata
3961 catacagata atggcagtaa cttcaccagt aatgcagtta aagcagcctg ttggtgggca
4021 ggtattcaac aggaatttgg aattccctac aatccccaaa gtcaaggagt agtagaatct
4081 atgaataaag aattaaagaa aatcataggg caggtaagag aacaagctga gcaccttaag
4141 acagcagtac aaatggcagt attcattcac aattttaaaa gaagaggggg gattgggggg
4201 tacagtgcag gggaaagaat aatagatata atagcatcag acatacagac taaagaactc
4261 caaaaccaaa ttttaaaaat tcaaaatttt cgggtttatt acagagacag cagagaccct
                                                    /\ 3' sj
4321 atttggaaag gaccagccaa actactctgg aaaggtgaag gggcagtagt aatacaagat
                                       /\ 5' sj
4381 aatagtgaca taaaggtagt accaaggagg aaagcaaaaa tcattaggga ttatggaaaa
                                                 vif cds start ->
4441 cagatggcag gtgctgattg tgtggcaggt agacaggatg aagattagaa catggaatag
                                                       <- pol cds
                                                          integrase end
4501 tttagtaaag caccatatgc atatttcaag gagagctaat ggatgggttt atagacatca
4561 ttatgacagc agacatccaa aggtaagttc agaagtacac atcccattag gggaggctag
4621 attaataata aaaacatatt ggggtttgca aacaggagaa agagattggc atttgggtca
4681 tggagtctcc atagaatgga gattgagaag ctataacaca caagtagacc ctggcctggc
4741 agaccaccta attcatatgc attattttga ttgttttgca gaatctgcca taaggaaagc
                                                    /\ 3' sj
4801 catattagga tatagagtta gccctaggtg tgactatcaa gcaggacata ataaggtagg
                                                                /\ 5' sj
4861 atctctacaa tacttggcac tgacagcatt gataaagcca aaaaaggcaa agccacctct
4921 gcctagtgtt agtaaattag tagaggataa atggaacaag ccccagaaga ccaggggccg
                        vpr cds start ->
4981 cagagggaac catacaatga atgggcacta gagcttttag aggagctcaa gcaggaagct
                                    < - vif cds end
5041 gtcagacact ttcctagacc atggctccat aacttaggac aatatatcta tgaaacctat
5101 ggggatactt ggtcgggagt agaagcttta ataagaactc tgcaacaact aatgtttatc
5161 catttcagaa ttggatgcca gcatagcaga ataggcattt tacgacagag aagagcaaga
           /\ 3' sj
5221 aatggagcca gtagatccta acctagaacc ctggaaccat ccaggaagtc agcctaagac
  tat ->                < - vpr cds end
start cds
5281 tgcttgtaat caatgttatt gtaaaaaatg tagctatcat tgtctagttt gctttctgac
5341 aaaggcttag gcatttccta tggcaggaag aagcggcgac agcgacgaag agctcctcaa
       /\ rev cds start - >    /\ 3' sj
  1 pt deletion
            <- tat cds premature end
               due to deletion
5401 agcagtaagg atcatcaaaa tcttatatca aagcagtaag taccaaataa tagatgtaat
                   (tat, rev, nef) 5' sj /\          vpu cds start ->
5461 ggttgattta ctagcaaaag tagattatag aatagtaata gtagcattca tagtagcact
5521 aatcatagca atagttgtgt ggaccatagc atatatagaa tataggaaat tgttaagaca
```

```
5581 aagaagaata gataggttaa ttaaaagaac tagggaaaga gcagaagaca gtggcaatga
                                                          env cds ->
                                                    signal peptide
5641 aagtgatggg gatacagagg aattgtcaac aatggtggat atggggaatc ttaggctttt
5701 ggatgttaat gatttgtaat ggaatgggga acttgtgggt cacagtctat tatggggtac
                     <- vpu cds end
      env signal peptide/ \gp120
5761 ctgtatggaa agatgctagc cctactctat tttgtgcatc agatgctaaa gcatatgata
5821 cagaagtaca taatgtctgg ggtacatttg cctgtgtccc cacagacccc agcccacaag
5881 aattgggttt ggaaaatgtg acagaaaatt ttaacatgtg gaaaaatgac atggtggagc
5941 agatgcatca ggatataatc agtttatggg atcaaggcct aaagccatgt gtaaagttga
6001 ccccactctg tgttacttta aactgtaatg ctatcaaaaa caatactaag gttacaaata
6061 atagtatcaa tagtgccaat gatgaaatga aaaattgttc tttcaatata accacagaac
6121 taagagataa gaaaaggaag gcatatgcac ttttttataa acttgatata gtaccactta
6181 ataacggttc tactgattat agattaataa attgtaatac ctcaaccata acacaagcct
6241 gtccaaaggt ctctttggac ccaattccta tacattattg tgctccagct ggttatgcga
6301 ttctaaagtg tagagataag acattcactg gaacaggacc atgccataat gtcagcacag
6361 tacaatgcac acatggaatt aaaccagtgg tatcaactca actattgtta aatggtagta
6421 tagcagaagg tgagacaata attagatttg aaaatctgac aaacaatgcc aaaataataa
6481 tagtacagct taatgaatct gtagaaatta cttgtacgag acccagcaat aatacaagag
                          V3 loop start ->
6541 aaagtataag gataggacca ggacaaacat tctatgcaac aggagacata ataggagata
6601 taagacaagc acattgtaac attagtgaag aaaaatggaa taaaactcta caaaaggtaa
                     <- V3 loop end
6661 aggaaaaatt acaaaagcac ttccctaata aaacaataga atttaagcca tcctcaggag
6721 gggacctaga aattacaaca catagcttta attgtggagg agaatttttc tattgcaata
6781 catcaaacct gtttaatagc acaaaactag aactgtttaa tagcagtaca aatttaaaca
6841 tcacactcca atgcagaata aaacaaatta taaacatgtg gcagggagta ggacgagcaa
6901 tgtatgcccc tcccattgaa ggaataataa tgtgtagatc aaatatcaca ggactactac
6961 tgacacgtga tggagccaaa gagccacata gcacaaaaga gatattcaga cctgaaggag
7021 gagatatgag ggacaattgg agaagtgaat tatataaata taaagtggta gaaattaagc
7081 cactaggagt agcacccact aagccaaaaa ggagagtggt ggagagagaa aaaagagcag
                                                          env gp120/\env gp41
7141 cactaggagc tttgttcctt gggttcttgg gagcagcagg aagcactatg ggcgcagcat
7201 caataacgct gacggtacag gccagacaat tattgtcggg tatagtgcaa cagcaaagca
7261 atttgctgaa agctatagag gcgcaacagc atatgttgca gctcacggtc tggggcatta
7321 agcagctcca gacaagagtc ctggctatag aaagacacct aagggatcaa cagctcctag
7381 ggatttgggg ctgctctgga aaactcatct gcaccactgc tgtgccttgg aactctagtt
7441 ggagtaataa gtctcaagag gagatttggg ataacatgac ctggatgcag tgggatagag
7501 aaattagtaa ttatacagac ataatataca atttgcttga agtctcgcaa aaccagcagg
7561 acaaaaatga aaaagattta ctagcattgg acaaatggga aaatctgtgg aattggttta
7621 acataacaaa ttggctgtgg tatataaaaa tattcataat gatagtggga ggtgtgatag
7681 gtttaagaat aatttttgct gtgctttcta tagtgaatag agttaggcag ggatactcac
7741 cattatcgtt tcagaccctt atcccacacc cgaggggacc cgacaggctc ggaggaatcg
                   /\ (tat, rev, nef) 3' sj
7801 aagaagaagg tggagagcaa ggcagagaca gatccatacg cttagtgaac ggattcttag
                                                          <- tat cds end
7861 caatttttg ggacgacctg cggagcctgt gcctcttcag ttaccaccgc ttgagagact
7921 taatcttgat tgcagcgagg acagtggaac ttctgggacg cagcagtctc aagggactac
7981 agaggggtg ggaaaccctt aaatatctgg gaagccttgt gcagtattgg ggtctggagc
                          <- premature rev end
8041 taaaaaagag tgctattaat ctgcttaata ccacagcaat agtagtaggt gaaggaacag
            <- rev cds normal end
8101 atagatttat agaattaata caaagaattt ggagagcttt ctgcaacata cctagaagaa
8161 taagacaggg cttggaagca gctttgcaat aaaatggggg gcacgatgtc aaaatgtagt
                                    <- env cds
                                       gp41 end
                          nef cds start ->
8221 ccagtaggat ggcctgctat aagagaaaga ataagacgag ctgctccagc agcagaggga
```

```
8281 gtaggagcag cgtctcgaga cttagacaaa tatggggcac ttacaagcag caacacaccc
8341 gccaataatc ctgattgtgc ctggctggaa gcgcaagagg aggaagaaga ggtaggcttt
8401 ccagtcagac ctcaggtacc tttaagacca atgacttata aggcagcatt cgatctcagc
8461 ctctttttaa aagaaaaggg gggactggaa gggttaattt actccaagaa aaggcaggag
                     3' LTR U3 ->
8521 attcttgatt tgtgggtcta taacacacaa ggcttcttcc ctgattggca aaactacaca
8581 ccaggaccag gggtcagata tccactgacc tttggatggt gcttcaagct agtaccagtt
8641 gacccaagtg aagtggaaga aatcaatgaa ggagagaaca actgcttgct gcaccctgcg
8701 agcctgcatg gaatggagga tgaagacaga gaagtattaa agtggaagtt tgacagtcac
8761 ctagcacgca gacacatggc ccgcgagcta catccggagt attacaaaga ctgctgacac
                                                              <- nef cds
                                                                 end
8821 agacgggact ttccgccggg actttccact ggggcgttcc aggaggaggg gtctgggcgg
        ------ ----   --- -------    ----------
      NFkB site III   NFkB site II   NFkB site I
8881 gactgggagt ggccaaccct cagatgctgc atataagcag ctgcttttcg cttgtaccgg
                                                            LTR U3/\R rpt
8941 gtctctctag gtagaccaga tctgagcctg ggagctctct ggctatctgg ggaacccact
9001 gcttaagcct caataaagct tgccttgagt g
               ------
           poly A signal
```

```
LOCUS       HIVU51188    9843 bp    DNA             VRL       18-JUL-1996
DEFINITION  Human immunodeficiency virus type 1 (HIV-1), clone 90cr402 from the
            Central African Republic, complete genome.
ACCESSION   U51188
KEYWORDS    .
SOURCE      Human immunodeficiency virus type 1 (HIV-1), clone 90cr402; PBMCs
            extracted from symptomatic male patient from Central African
            Republic; expanded in chimpanzee cells and re-expanded in human
            PBMCs; complete genome lambda cloned and sequenced.
  ORGANISM  Human immunodeficiency virus type 1
            Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;
            Retroviridae; Lentivirinae.
REFERENCE   1  (bases 1 to 9843)
  AUTHORS   Gao,F., Morrison,S.G., Robertson,D.L., Thornton,C.L., Craig,S.,
            Karlsson,G., Sodroski,J., Morgado,M., Galvao-Castro,B., von
            Briesen,H., Beddows,S., Weber,J., Sharp,P.M., Shaw,G.M. and
            Hahn,B.H.
  TITLE     Molecular cloning and analysis of functional envelope genes from
            human immunodeficiency virus type 1 sequence subtypes A through G.
            The WHO and NIAID Networks for HIV Isolation and Characterization
  JOURNAL   J. Virol. 70 (3), 1651-1657 (1996)
  MEDLINE   96190564
REFERENCE   2  (bases 1 to 9843)
  AUTHORS   Gao,F., Robertson,D.L., Morrison,S.G., Hui,H., Craig,S.,
            Fultz,P.N., Decker,J., Girard,M., Shaw,G.M., Hahn,B.H. and
            Sharp,P.M.
  TITLE     The heterosexual HIV-1 epidemic in Thailand is caused by an
            intersubtype (A/E) recombinant of African origin
  JOURNAL   J. Virology 70 (10), 7013-7029 (1996)
REFERENCE   3  (bases 1 to 9843)
  AUTHORS   Gao,F., Robertson,D.L., Morrison,S.G., Hui,H., Craig,S.,
            Fultz,P.N., Decker,J., Girard,M., Shaw,G.M., Hahn,B.H. and
            Sharp,P.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-1996) John Blouin, HIV Database, Los Alamos
            National Laboratory, Eniwetok Dr., Los Alamos, NM 87545, USA
  COMMENT   A set of three complete genomes from a study linking the HIV-1
            epidemic in the heterosexual population in Thailand to an A/E
            recombinant. 90cr402, previously named CAR-E 4002, was obtained
            from a man from Bangui, Central African Republic, who had
            lymphadeopathy, diarrhea, severe weight loss and recurrent
            respiratory infections. He was infected through heterosexual
            contact, but the year of infection is unknown. 90cr402 was first
            adapted to growth in chimpanzee cells, expanded in chimpanzee
            cells, and then re-expanded in human PBMCs before lamba cloning
            and sequencing. 90cr402 and another sequence in the study, 93th253,
            are subtype A/E recombinants, and comparison of the two strains
            shows that they were derived from a common A/E recombinant ancestor,
            presumably from Central Africa.
FEATURES             Location/Qualifiers
     source          1..9843
                     /organism="Human immunodeficiency virus type 1"
                     /note="from the Central African Republic"
                     /clone="90cr402"
     LTR             1..656
     CDS             832..2328
                     /gene="gag"
                     /codon_start=1
                     /product="gag polyprotein"
```

```
                /translation="MGARASILSGGKLDAWEKIRLRPGGKKQYRMKHLVWASRELER
                FALNPGLLETAEGCQQLIEQLQSTIKTGSEELKSLFNTIATLWCVHQRIDVKDTKEA
                LDKIEEVQNKSKQKAQQAAAGTGSSSKVSQNYPIVQNAQGQMVHQALSPRTLNAWVK
                VVEEKGFNPEVIPMFSALSEGATPQDLNMMLNIVGGHQAAMQILKDTINEEAAEWDR
                VHPVHAGPIPPGQMREPRGADIAGTTSTLHEQIGWMTSNPPIPVGEIYKKWIILGLN
                KIVRMYSPVSILDIRQGPKEPFRDYVDRFFKTLRAEQATQEVKNWMTETLLVQNANP
                DCKSILKALGTGATLEEMMTACQGVGGPGHKARVLAEAMSQVHHTNIMMQKGNFKGQ
                RKIKCFNCGKEGHLARNCRAPRKKGCWKCGREGHQMKDCTERQANFLGKIWPSSKGR
                PGNFPQSRPEPTAPPMESLGMGEEITSFPKQEQKDKKQPPPLVSLKSLFGNDPLSQ"
CDS             join(832..2120,2120..5132)
                /codon_start=1
                /product="gag-pol polyprotein"
                /translation="MGARASILSGGKLDAWEKIRLRPGGKKQYRMKHLVWASRELER
                FALNPGLLETAEGCQQLIEQLQSTIKTGSEELKSLFNTIATLWCVHQRIDVKDTKEA
                LDKIEEVQNKSKQKAQQAAAGTGSSSKVSQNYPIVQNAQGQMVHQALSPRTLNAWVK
                VVEEKGFNPEVIPMFSALSEGATPQDLNMMLNIVGGHQAAMQILKDTINEEAAEWDR
                VHPVHAGPIPPGQMREPRGADIAGTTSTLHEQIGWMTSNPPIPVGEIYKKWIILGLN
                KIVRMYSPVSILDIRQGPKEPFRDYVDRFFKTLRAEQATQEVKNWMTETLLVQNANP
                DCKSILKALGTGATLEEMMTACQGVGGPGHKARVLAEAMSQVHHTNIMMQKGNFKGQ
                RKIKCFNCGKEGHLARNCRAPRKKGCWKCGREGHQMKDCTERQANFFRENLAFQQGE
                ARKFPSEQTRANSPTNGELGDGGRDNLLPEAGAERQETASSFSFPQITLWQRPLVTV
                KVGGQLKEALLDTGADDTVLEDINLPGKWKPKMIGGIGGFIKVRQYDQILIEICGKK
                AIGTVLVGPTPVNIIGRNMLTQIGCTLNFPISPIDTVPVTLKPGMDGPKVKQWPLTE
                EKIKALTEICKEMEEEGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQ
                DFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEGFRKYTAFTIPSINNETPGV
                RYQYNVLPQGWKGSPAIFQSSMTKILEPFRARNPEIVIYQYMDDLYVGSDLEIGQHR
                TKVEDLRAHLLSWGFTTPDKKHQKEPPFLWMGYELHPDRWTVQPIVLPEKDSWTVND
                IQKLVGKLNWASQIYAGIKVRQLCKLLRGAKALTDIVTLTEEAELELAENREILKTP
                VHGVYYDPSKDLVAEIQKQGQDQWTYQIYQEPFKNLKTGKYARKRSAHTNDVRQLAE
                VVQKVATESIVIWGKTPKFRLPIQRETWETWWMEYWQATWIPEWEFVNTPPLVKLWY
                QLEKDPIMGAETFYVDGAASRETKQGKAGYVTDRGRQKVVSLTETTNQKTELHAIHL
                ALQDSGPEVNIVTDSQYALGIIQAQPDRSESDIVNQIIEKLIEKEKVYLSWVPAHKG
                IGGNEQVDKLVSSGIRKVLFLDGIDKAQEDHERYHSNWRAMASDFNLPPIVAKEIVA
                SCDKCQLKGEAMHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQ
                ETAYFLLKLAGRWPVRVIHTDNGSNFTSAAVKAACWWANVQQEFGIPYNPQSQGVVE
                SMNKELKKIIGQVREQAEHLKTAVQMAVFIHNFKRKGGIGEYSAGERIIDIIATDIQ
                TKALQKQITKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAVVIQDNSDIKVVPRRKAK
                IIRDYGKQMAGDDCVAGRQNED"
CDS             5077..5355
                /gene="vif"
                /note="premature end at 5355. Normal end would be at 5655."
                /codon_start=1
                /product="vif protein"
                /translation="MENRWQVMIVWQVDRMRIRAWNSLVKHHMYSSKKAAKWFYRHHY
                ESQHPKVSSEVHIPLGDARLIIRTYWGLHTGEKDWHLGHGVSIEWRQRK"
CDS             5084..5374
                /gene="vpr"
                /codon_start=1
                /product="vpr protein"
                /translation="MEQAPEDQGPQREPYNEWTLELLEELKTEAVRHFPRPWLHGLGK
                HIYNTYGDTWEGVEAIIRILQQLLFVHFRIGCQHSRIGIIRGTRGRNGAGRS"
CDS             join(5866..6080,8442..8526)
                /gene="tat"
                /codon_start=1
                /product="tat protein"
                /translation="MEPVDPNLEPWKHPGSQPTTACSKCYCKMCCWHCQLCFLKKGLG
                ISYGRKKRKHRRGPSQDSKDHQNSIPKQPLPTSRGNPTGPKESKKKVESKAETDP"
```

```
     CDS             join(6005..6080,8442..8737)
                     /gene="rev"
                     /codon_start=1
                     /product="rev protein"
                     /translation="MAGRSGSTDEDLLRTVRIIKILYQSNPYPPAEGTRQARKNRRRK
                     WRARQRQIHKIGERILSTCLGRSPEPVPLQLPPLERLHLDCSEDCGTSGTQQSTGTET
                     EVGRPQISGESSVILGSGTKN"
     CDS             6101..6346
                     /gene="vpu"
                     /codon_start=1
                     /product="vpu protein"
                     /translation="RLPLHICAIVGLIVALIIAIVVWTIVAIEYKKLRRQRKIDRLVQ
                     RISERAEDSGNESEGDTEELAKLVEMGDFDPWVGDNL"
     CDS             6264..8879
                     /gene="env"
                     /codon_start=1
                     /product="env polyprotein"
                     /translation="MRVKGTRRNWPNLWKWGTLILGLVIICSASDNLWVTVYYGVPVW
                     RDADTILFCASDAKAHVTEVHNVWATHACVPTDPNPQEIYLENVTENFDMWKNNMVEQ
                     MQEDVISLWDQSLQPCVKLTPLCVTLHCTKASFTNATSDRIKMEDAVRNCSFNMTTEL
                     QDKKQEVHALFYTSDVVQISSSVQNNNNSNTSGQNNSHKFRLIHCNTSVIKQACPKVS
                     FDPIPIHYCAPAGYAILKCNDKNFNGTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAE
                     EEIIIRSEDLTDNAKTIIVHLNKSIEINCTRPFKKVRISARIGPGRVFHTTGNINGDI
                     RKAYCEINKTKWKETLKQVTRKLREHLNGTMTISFRPSSGGDPEITMHHFNCRGEFFY
                     CNTTALFNSTWINGTMQEVNGTNSGNITLPCRIKQIVNMWQEVGRAMYAPPISEVINC
                     VSNITGILLTRDGGINQNQTNKNETFRPGGGNIKDNWRSELYKYKVVQIEPLGIAPTK
                     ARRRVVEREKRAVGIGAMIFGFLGAAGSTMGAASITLTVQARQLLSGIVQQQSNLLRA
                     IEAQQHLLQLTVWGIKQLQARVLAVERYLKDQKFLGLWGCSGKIICTTSVPWNSSWSN
                     KSYEAIWNNMTWIEWDREINEYTNQIYELLTESQDQQERNEKDLLALDKWASLWNWFD
                     ITRWLWYIKIFIIIVGGLIGLRIVFAVLSIVNRVRQGYSPLSFQTLTHQQREPDRPER
                     IEEESGEQGRDRSIRLVSGFLALAWDDLRSLCLFSYHRLRDFILIAARTVELLGHSSL
                     QGLRRRWEGLKYLGNLLSYWVQELRISAITLLDATAITVAGWTDRVIEIVQRAWRAIL
                     HIPRRIRQGLERALL"
     CDS             8881..9516
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /translation="MGGKWSKSCIVGWPQVRERIRQTPVAEERQTPAAAEGVGAVSQD
                     LDKHGAVTSSNINNADNVWLEAQEEEEVGFPVRPQVPLRPMTYKGAFDLSFFLKEKGG
                     LDGLIHSKRRQEILDLWVHNTQGYFPDWQNYTPGPGVRYPLCFGWCFKLVPVDPREVE
                     EDNKGENNCLLHPMSQHGIDDDEREVLMWKFDSSLARRHIARELHPEYYKD"
     LTR             9188..9843
     misc_signal     386..395
                     /note="NFkappaB"
     misc_signal     9573..9582
                     /note="NFkappaB"
BASE COUNT     3501 a   1764 c   2400 g   2178 t
ORIGIN
        1 tggatgggct aattcactcc aagagaagac aagagatcct tgacttatgg gtccacaata
5' LTR U3 ->
       61 cacaaggcta cttccctgat tggcaaaact acacaccagg gccgggggtc agatacccac
      121 tgtgttttgg atggtgcttc aagctagtac cagttgatcc aagagaagta gaggaggaca
      181 ataaaggaga aaacaactgc ctgttacacc ccatgagcca gcatggaata gatgatgatg
      241 aaagagaagt gctaatgtgg aagtttgaca gttccctagc acgaagacac atagctcgag
      301 agctgcaccc agagtactac aaagactgat gacaaagaag ttccaagact gctgacaaag
                                                                  <- nef
                                                                     cds end
      361 gagtttctaa ctaggacttc cgctggggac tttccagggg aggtgtggcc tgggcggggt
                                  ------ ------ Nf Kappa B
```

```
 421 tggggagtgg ctaaccctca gatgctgcat aaaagcagcc gcttttcgct tgtactgggt
                                                          LTR U3/\R rpt
 481 ctctcttgtt agaccaggtc gagcccggga gctctctggc tagcagggga acccactgct
 541 taaagcctca ataaagcttg ccttgagtgc ttaaagtggt gtgtgcccgt ctgtgttagg
             - -----               R rpt/\LTR U5
          poly-A signal
 601 actctggtaa ctagagatcc ctcagaccac tctagactga gtaaaaatct ctagcagtgg
                                                            <- LTR
 661 cgcccgaaca gggacttgaa agcgaaagtt aatagggact cgaaagcgaa agttccagag
 721 aagttctctc gacgcaggac tcggcttgct gaggtgcaca cagcaagagg cgagagcggc
 781 gactggtgag tacgccaaaa attttgacta gcggaggcta gaaggagaga gatgggtgcg
       /\ major 5' sj                              gag cds ->
                                                 p17 MA start
 841 agagcgtcaa tattaagtgg gggaaaatta gatgcatggg agaaaattcg gctaaggcca
 901 gggggaaaga aacaatatag aatgaaacat ttagtatggg caagcagaga gttagaaaga
 961 ttcgcactta accctggcct tttagaaaca gcagagggat gtcaacaact aatagaacag
1021 ttacaatcaa ctatcaagac aggatcagaa gaacttaaat cactatttaa tacaatagca
1081 actctctggt gcgtacatca aaggatagat gtaaaagaca ccaaggaagc tttagataaa
1141 atagaggaag tacaaaataa gagcaagcaa aaagcacagc aggcagcagc tggcacagga
1201 agcagcagca aagtcagcca aaattaccct atagtgcaaa atgcacaagg gcaaatggta
                       gag p17 MA/\gag p24 CA
1261 catcaggcct tatcacctag aactttgaat gcatgggtga aagtagtaga agagaagggt
1321 tttaacccag aagtaatacc catgttctca gcattatcag aaggagccac cccacaagat
1381 ttaaatatga tgctcaacat agtgggggga caccaggcag caatgcaaat attaaaagat
1441 accatcaatg aggaagctgc agaatgggat agggtacacc cagtacatgc agggcctatt
1501 ccaccaggcc agatgagaga accaagggga gctgacatag caggaactac tagtaccctt
1561 catgaacaga taggatggat gacaagcaac ccacctattc cagtgggaga aatctataaa
1621 aagtggataa tcctgggatt aaataaaata gtaagaatgt atagccctgt tagcattttg
1681 gacataagac aagggccaaa agaacccttc agagactatg tagacaggtt ctttaaaact
1741 ctcagagcgg aacaagctac acaggaggta aaaaactgga tgacagaaac tttgctagtc
1801 caaaatgcga acccagactg taagtccatt ttaaaagcac taggaacagg ggctacatta
1861 gaagaaatga tgacagcatg ccagggagtg ggaggacctg gacataaagc aagggttttg
                                                       gag p24 CA/
1921 gctgaggcaa tgagtcaagt acaccataca aatataatga tgcagaaagg caattttaag
     \gag p2x                          gag p2x/\p7NC
                                              \gag-pol TF
1981 ggccagagaa aaattaagtg tttcaactgt ggcaaagaag gacacctagc cagaaattgc
2041 agggccccta gaaaaaaggg ttgttggaaa tgtgggaggg aaggacacca aatgaaagac
2101 tgcactgaaa gacaggctaa ttttttaggg aaaatctggc cttccagcaa ggggaggcca
                           ------          ^^^^^ ^^ ^^       ^^ ^^^^^^
       gag-pol ribosomal frameshift site      stem      loop     stem
2161 ggaaatttcc ctcagagcag accagagcca acagccccac caatggagag cttggggatg
     ^
     gag p7NC/\gag P6
2221 ggggaagaga taacctcctt cccgaagcag gagcagaaag acaagaaaca gcctcctcct
2281 ttagtttccc tcaaatcact ctttggcaac gacccttgt cacagtaaaa gtaggaggac
  gag-pol TF/\pol protease                              <- gag cds
                                                          P6 end
2341 agctaaaaga agctctatta gacacaggag cagatgatac agtattagaa gacataaatt
2401 tgccaggaaa atggaaacca aaaatgatag ggggaattgg aggttttatc aaagtaagac
2461 aatatgatca gatacttata gaaatttgtg ggaaaaaggc tataggtaca gtgttagtag
2521 gacctacccc cgtcaacata attggacgga acatgttgac tcagattggt tgtactttaa
2581 atttcccaat tagtcctatt gacactgtac cagtaacatt aaagccagga atggatggac
pol protease/\pol RT
2641 caaaggttaa acaatggcca ttgacagaag aaaaaataaa agcattaaca gaaatttgta
2701 aagaaatgga agaggaagga aaaatctcaa aaattgggcc tgaaaatcca tacaatactc
2761 cagtatttgc tataaagaaa aaggacagca ctaaatggag aaaattagta gatttcagag
2821 agctcaataa aagaactcag gacttttggg aagttcaatt aggaatacca catccagcag
2881 gtctaaaaaa gaaaaaatca gtaacagtac tagatgtggg agacgcatat ttttcagttc
```

```
2941 ctttagatga aggctttaga aagtatactg cattcaccat acctagtata aacaatgaga
3001 caccaggagt cagatatcag tacaatgtgc tgccgcaggg atggaaagga tcaccagcaa
3061 tattccagag tagcatgaca aaaatcttag agccctttag agcaagaaat ccagaaatag
3121 ttatctatca atacatggat gacttgtatg taggatctga tttagaaata gggcagcata
3181 gaacaaaagt agaggatcta agagctcatc tattgagctg gggatttact acaccagaca
3241 aaaagcatca gaaagaaccg ccattccttt ggatgggata tgaactccat cctgacagat
3301 ggacagtcca gcctatagtg ctgccagaaa aagacagctg gactgtcaat gatatacaga
3361 agttagtggg aaaactaaat tgggcaagtc aaatttatgc agggattaaa gtgaggcaac
3421 tgtgtaaact cctcagggga gctaaagcac taacagacat agtaacactg actgaggaag
3481 cagaattaga attggcagag aacagggaaa ttctaaaaac ccctgtgcat ggggtatatt
3541 atgacccatc aaaagactta gtagcagaaa tacagaaaca agggcaagac caatggacat
3601 atcaaattta tcaagagcca tttaaaaatc taaaaacagg aaaatatgca agaaaaaggt
3661 ctgctcacac taatgatgta agacaattag cagaagtggt gcaaaaagtg gccacagaaa
3721 gcatagtaat atggggaaag acccctaaat ttagactacc catacaaaga gaaacatggg
3781 agacatggtg gatggagtat tggcaggcta cctggattcc tgaatgggag tttgtcaata
3841 cccctcctct agtaaaatta tggtaccaat tagaaaaaga ccccataatg ggagcagaga
3901 ctttctatgt agatggggca gctagtaggg agactaagca aggaaaagca gggtatgtca
   pol RT/\pol RNAse
3961 ctgacagagg aagacaaaaa gtagtttccc taactgagac aacaaatcaa aagactgaat
4021 tacatgcgat ccatttagcc ttgcaggatt caggaccaga agtaaacata gtaacagact
4081 cacaatatgc attaggaatt attcaggcac aaccagacag gagtgagtca gatatagtca
4141 atcaaataat agagaaacta atagaaaagg aaaaagtcta cctgtcatgg gtcccagcac
4201 ataaggggat tggaggaaat gaacaagtag ataaattagt cagttctgga atcaggaagg
4261 tgctattttt agatgggata gataaggctc aagaagacca tgaaagatat cacagcaatt
pol RNAse/\pol integrase
4321 ggagagcaat ggctagtgac tttaatttgc cacctatagt agcaaaggaa atagtagcca
4381 gctgtgataa atgtcagcta aaaggggaag ccatgcatgg acaagtagac tgtagtccag
4441 ggatatggca actagattgc acgcatctag aaggaaaagt catcctggta gcagtccacg
4501 tggccagtgg atatatagaa gcagaagtta tcccagcaga aacaggacag gagacagcat
4561 actttctgct aaaattagca ggaagatggc cagtaagggt aatacacaca gacaatggca
4621 gcaatttcac cagcgctgca gttaaagcag cctgttggtg ggccaatgtc caacaggaat
4681 ttggaattcc ctacaatccc caaagccaag gagtagtgga atctatgaat aaggaattaa
4741 agaaaatcat agggcaggta agagagcaag ctgaacacct taagacagca gtacaaatgg
4801 cagtattcat tcacaatttt aaaagaaaag gggggattgg ggagtacagt gcaggggaaa
4861 gaataataga cataatagca acagacatac aaactaaagc attacaaaaa caaattacaa
4921 aaattcaaaa ttttcgggtt tattacaggg acagcagaga tccaatttgg aaaggaccag
4981 caaaactact ctggaaaggt gaaggggcag tagtaataca ggacaatagt gatataaaag
5041 tagtaccaag aagaaaagca aagatcatta gggattatgg aaaacagatg gcaggtgatg
                              vif cds start ->
5101 attgtgtggc aggtagacag aatgaggatt agagcatgga acagtttagt aaaacatcat
                                     <- pol cds
                                        integrase end
5161 atgtatagct caaagaaagc tgccaaatgg ttttatagac atcattatga aagccagcat
5221 ccaaaagtaa gttcagaagt acacatccca ctaggggatg ctagattaat aataagaaca
5281 tattggggtc tgcatacagg agaaaaagac tggcacttgg gtcatggagt ctccatagaa
5341 tggaggcaga gaaaatagag cacacaaata gatcctgacc tagcagacca actgattcat
                  <- vif premature end
5401 ctgcaatatt ttgactgttt ttcagactct gccataagga aagccttatt aggacaaata
                                /\ 3' sj
5461 gttagaccta ggtgtgaata tccagcagga cataacaagg taggatctct acaatatttg
                                                 /\ 5' sj
5521 gcactgaaag cattaacaaa aacaaaaaag acaaagccac ctctgcctag tgttaggaaa
5581 ttaacagaag atagatggaa caagccccag aagaccaagg gccacagaga gagccctaca
      vpr cds start ->
5641 atgaatggac attagaactg ttagaggagc ttaaaactga agctgttaga cattttccta
                  <- vif cds end
5701 ggccctggct ccatggctta ggaaagcata tctataacac ttatggggat acttgggaag
5761 gggttgaagc tataataaga attttgcaac aactactgtt tgttcatttc agaattgggt
```

```
    5821 gccaacatag cagaataggc attatacgag ggacaagagg caggaatgga gccggtagat
                                          tat cds start ->          <- vpr
                                                                       cds end
    5881 cctaacctag agccctggaa acatccggga agtcagccta caactgcttg tagcaagtgt
    5941 tattgtaaaa tgtgttgctg gcattgccaa ctatgctttc tgaaaaaagg cttaggcatc
    6001 tcctatggca ggaagaagcg gaagcaccga cgaggacctt ctcaggacag taaggatcat
rev cds start->
    6061 caaaattcta taccaaagca gtaagtatta agtatatata aggttgcctt tgcatatctg
             tat-rev exon/ \intron
                           vpu cds normal start ->
                           ATG in others, AGG here
    6121 tgcaatagta ggactgatag tagcgctaat catagcaata gtagtgtgga ctatagtggc
    6181 tatagaatat aagaaattaa gaaggcaaag aaaaatagat aggttagttc aaagaataag
    6241 tgaaagagca gaagacagtg gaaatgagag tgaaggggac acggaggaat tggccaaact
                        env cds ->
         signal peptide start
    6301 tgtggaaatg ggggactttg atccttgggt tggtgataat ttgtagtgcc tcagacaact
                                                    <- vpu cds end
                                         env signal pept/\gp120
    6361 tgtgggttac agtttattat ggggttcctg tgtggagaga tgcagatacc atcctatttt
    6421 gtgcatcaga tgccaaagca catgtgacag aagtgcacaa tgtttgggct acacatgcct
    6481 gtgtacccac agaccccaac ccacaagaaa tatacctgga aaatgtaaca gaaaattttg
    6541 acatgtggaa aaataacatg gtagagcaga tgcaggagga tgtaatcagt ttatgggatc
    6601 aaagtctaca gccatgtgta aagttaactc ctctctgtgt tactttacat tgtaccaagg
    6661 ctagttttac taatgccacc agtgacagaa taaaaatgga agatgcagta agaaactgtt
    6721 cttttaatat gaccacagaa ctacaagata aaaagcagga ggtccatgca ctttttttata
    6781 cgtctgatgt agtacaaatt agtagtagtg tacaaaataa taataacagt aatactagtg
    6841 gacaaaataa tagtcataag tttagattaa tacattgtaa tacttcagtc attaagcagg
    6901 cttgtccaaa ggtatccttt gatccaattc ccatacatta ttgtgctcca gctggttatg
    6961 cgattctaaa gtgtaatgat aagaatttta atggaacagg gccatgcaaa aatgtcagct
    7021 cagtacaatg cacacatgga attaagccag tggtatcaac tcaattgctg ttaaatggca
    7081 gtctagcaga agaagagata ataatcagat ctgaagatct cacagacaat gccaaaacca
    7141 taatagtgca ccttaataaa tctatagaaa tcaattgtac cagacccttc aagaaagtaa
    7201 gaataagtgc aaggatagga ccaggacgag tattccatac aacaggaaac ataaatggtg
    7261 atataagaaa agcatattgt gaaattaata aaacaaaatg gaaagaaact ttaaaacagg
    7321 taacaagaaa attaagagag caccttaatg ggacaatgac aataagcttt cgaccatcct
    7381 caggaggaga tccagaaatc acaatgcatc attttaattg tagaggggaa tttttctatt
    7441 gcaatacaac agcactgttt aatagtactt ggataaatgg aaccatgcag gaggttaatg
    7501 gcacaaactc aggcaatatc acacttccat gcaggataaa gcaaattgta aacatgtggc
    7561 aggaagtagg acgagcaatg tatgctcctc ccatcagtga agtaattaat tgtgtatcaa
    7621 atattacagg aatactattg acaagagatg ggggtattaa tcaaaatcag actaacaaaa
    7681 atgagacctt cagacctgga ggaggaaata taaaggacaa ttggagaagt gaattataca
    7741 aatataaagt agtacaaatt gaaccactag gaatagcacc caccaaggca aggagaagag
    7801 tggtggagag agaaaaaaga gcagtgggaa taggagctat gatctttggg ttcttaggag
                     env gp120/ \gp41
    7861 cagcaggaag cactatgggc gcggcgtcaa taacgctgac ggtacaggcc agacaattat
    7921 tgtctggtat agtgcaacag caaagcaatt tgctgagggc tatagaggcg cagcagcatc
    7981 tgttgcaact cacagtctgg ggcattaaac agctccaggc aagagtcctg gctgtggaaa
    8041 gatacctaaa agatcaaaag ttcctaggac tttggggctg ctctggaaaa atcatctgca
    8101 ccacttctgt gccttggaac tccagttgga gtaataaatc ttatgaagct atttggaaca
    8161 acatgacatg gatagaatgg gatagagaaa ttaacgaata cacaaaccaa atatatgagc
    8221 tacttacaga atcgcaggac cagcaggaaa ggaatgaaaa ggatttgtta gcattggaca
    8281 agtgggcaag tctgtggaat tggtttgaca taacaaggtg gctgtggtat ataaaaatat
    8341 ttataataat agtaggaggt ttaataggtt taagaatagt ttttgctgtg ctttctatag
    8401 taaatagagt taggcaggga tactcacctt tgtcgttcca gacccttacc caccagcaga
                                      tat-rev intron/\exon
    8461 gggaacccga caggcccgaa agaatcgaag aagaaagtgg agagcaaggc agagacagat
    8521 ccataagatt ggtgagcgga ttcttagcac ttgcctggga cgatctccgg agcctgtgcc
             <- tat cds end
    8581 tcttcagcta ccaccgcttg agagacttca tcttgattgc agcgaggact gtggaacttc
    8641 tgggacacag cagtctacag ggactgagac ggaggtggga aggcctcaaa tatctgggga
```

```
   8701 atcttctgtc atattgggtt caggaactaa gaattagtgc tattactttg cttgatgcta
                                            <- rev cds end
   8761 cagcaataac agtagcgggg tggacagata gggttataga aatagtacaa agagcttgga
   8821 gagctatcct ccatatacct agaagaatca gacagggctt ggaaagggct ttgctataac
                                                                  <- env cds
                                                                     gp41 end
   8881 atgggtggca aatggtcaaa aagttgcata gtgggatggc ctcaggtcag ggaaagaata
nef cds ->
   8941 aggcaaactc ctgtagcaga agaaaggcaa actcctgcag cagcagaagg agtaggagca
   9001 gtatctcaag atctagataa acatggagca gtaacaagca gtaatataaa taatgctgat
   9061 aatgtctggc tggaagcgca agaggaagag gaggtaggct ttccagtcag gccgcaggta
   9121 cctctaagac caatgactta taagggagct tttgatctta gcttcttttt aaaagaaaag
   9181 gggggactgg atgggctaat tcactccaag agaagacaag agatccttga cttatgggtc
        LTR ->
   9241 cacaatacac aaggctactt ccctgattgg caaaactaca caccagggcc gggggtcaga
   9301 tacccactgt gttttggatg gtgcttcaag ctagtaccag ttgatccaag agaagtagag
   9361 gaggacaata aaggagaaaa caactgcctg ttacacccca tgagccagca tggaatagat
   9421 gatgatgaaa gagaagtgct aatgtggaag tttgacagtt ccctagcacg aagacacata
   9481 gctcgagagc tgcacccaga gtactacaaa gactgatgac aaagaagttc caagactgct
                                          <- nef cds end
   9541 gacaaaggag tttctaacta ggacttccgc tggggacttt ccaggggagg tgtggcctgg
                                         ---------- --
                                          nf kappa B
   9601 gcggggttgg ggagtggcta accctcagat gctgcataaa agcagccgct tttcgcttgt
   9661 actgggtctc tcttgttaga ccaggtcgag cccgggagct ctctggctag caggggaacc
         U3/\R rpt
   9721 cactgcttaa agcctcaata aagcttgcct tgagtgctta aagtggtgtg tgcccgtctg
                        ---- --               R rpt/\U5
                       poly-A signal
   9781 tgttaggact ctggtaacta gagatccctc agaccactct agactgagta aaaatctcta
   9841 gca
         <- 3' LTR U5 end
```

```
LOCUS       HIVU51190    8980 bp    DNA             VRL       18-JUL-1996
DEFINITION  HIV-1 clone 92ug037 from Uganda, complete genome
ACCESSION   U51190
KEYWORDS    .
SOURCE      Human immunodeficiency virus type 1 (HIV-1), clone 92ug037 from
            Uganda; genomic DNA isolated from PBMCs; genome extended by nested
            PCR amplification.
  ORGANISM  Human immunodeficiency virus type 1
            Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;
            Retroviridae; Lentivirinae.
REFERENCE   1  (bases 1 to 8980)
  AUTHORS   Gao,F., Morrison,S.G., Robertson,D.L., Thornton,C.L., Craig,S.,
            Karlsson,G., Sodroski,J., Morgado,M., Galvao-Castro,B., von
            Briesen,H., Beddows,S., Weber,J., Sharp,P.M., Shaw,G.M. and
            Hahn,B.H.
  TITLE     Molecular cloning and analysis of functional envelope genes from
            human immunodeficiency virus type 1 sequence subtypes A through G.
            The WHO and NIAID Networks for HIV Isolation and Characterization
  JOURNAL   J. Virol. 70 (3), 1651-1657 (1996)
  MEDLINE   96190564
REFERENCE   2  (bases 1 to 8980)
  AUTHORS   Gao,F., Robertson,D.L., Morrison,S.G., Hui,H., Craig,S.,
            Fultz,P.N., Decker,J., Girard,M., Shaw,G.M., Hahn,B.H. and
            Sharp,P.M.
  TITLE     The heterosexual HIV-1 epidemic in Thailand is caused by an
            intersubtype (A/E) recombinant of African origin
  JOURNAL   J. Virology 70 (10), 7013-7029 (1996)
REFERENCE   3  (bases 1 to 8980)
  AUTHORS   Gao,F., Robertson,D.L., Morrison,S.G., Hui,H., Craig,S.,
            Fultz,P.N., Decker,J., Girard,M., Shaw,G.M., Hahn,B.H. and
            Sharp,P.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-1996) John Blouin, HIV Database, Los Alamos
            National Laboratory, Enitwetok Dr., Los Alamos, NM 87545, USA
 COMMENT    A set of three complete genomes from a study linking the HIV-1
            epidemic in the heterosexual population in Thailand to an A/E
            recombinant. 92ug037 was obtained from WHO-NIAIDS and comes from
            an asymptomatic 31-year-old female from Entebbe, Uganda, early in
            infection. The year of infection is unknown and the mode of
            infection was heterosexual contact. 92ug037, U51190, was
            established and propagated by cocultivation with normal donor
            lymphocytes, and then PCR amplified and sequenced. 92ug037
            is subtype A.
FEATURES             Location/Qualifiers
     source          1..8980
                     /organism="Human immunodeficiency virus type 1"
                     /note="from Uganda"
                     /clone="92ug037"
     CDS             160..1659
                     /gene="gag"
                     /codon_start=1
                     /product="gag polyprotein"
                     /translation="MGARASVLSGGKLDAWEKIRLRPGGKKKYRLKHLVWASRELERF
                     ALNPSLLETTEGCQQIMEQLQSALRTGTEELRSLYNTVATLYCVHQRIEVKDTKEALD
                     KIEEIQKKSKQKTQQAAADTGSSSKVSQNYPIVQNAQGQMIHQSLSPRTLNAWVKVIE
                     EKALSPEVIPMFSALSEGATPQDLNMMLNIVGGHQAAMQMLKDTINEEAAEWDRLHPV
                     HAGPVAPGQMREPRGSDIAGTTSTPQEQIAWMTGNPPIPVGDIYKRWMILGLNKIVRM
                     YSPVSILDIKQGPKEPFRDYVDRFFKTLRAEQATQEVKGWMTETLLIQNANPDCKSIL
                     RALGAGATLEEMMTACQGVGGPGHKARVLAEAMSQVQHTNIMMQRGNFKGQKRIKCFN
                     CGKEGHLAKNCRAPRKKGCWKCGREGHQMKDCTERQANFLGKIWPSSKGRPGNFPQSR
                     PEPTAPPAAEIFGMREEIVSPPKQEQNDRDQNPPSVSLKSLFGNDLLSQ"
```

```
CDS             join(160..1450,1450..3146)
                /note="Premature stop codon at 3146, normal end would be
                at 1642."
                /codon_start=1
                /product="gag-pol polyprotein"
                /note="Gag-Pol fusion polpyprotein is produced by
                ribosomal frameshift slip on tttttt slippery site
                at 1450-1455 potentiated by an mRNA stem-loop structure
                at 1450-1455"
                /translation="MGARASVLSGGKLDAWEKIRLRPGGKKKYRLKHLVWASRELERFA
                LNPSLLETTEGCQQIMEQLQSALRTGTEELRSLYNTVATLYCVHQRIEVKDTKEALDKI
                EEIQKKSKQKTQQAAADTGSSSKVSQNYPIVQNAQGQMIHQSLSPRTLNAWVKVIEEKA
                LSPEVIPMFSALSEGATPQDLNMMLNIVGGHQAAMQMLKDTINEEAAEWDRLHPVHAGP
                VAPGQMREPRGSDIAGTTSTPQEQIAWMTGNPPIPVGDIYKRWMILGLNKIVRMYSPVS
                ILDIKQGPKEPFRDYVDRFFKTLRAEQATQEVKGWMTETLLIQNANPDCKSILRALGAG
                ATLEEMMTACQGVGGPGHKARVLAEAMSQVQHTNIMMQRGNFKGQKRIKCFNCGKEGHL
                AKNCRAPRKKGCWKCGREGHQMKDCTERQANFFRENLAFQQREARKFSSEQTRTNSPTS
                SRDLWDEGRDSLPSEAGAERQGPEPTFSFPQITLWQRPLVTVKIGGQLKKALLDTGADD
                TVLEDINLPGKWKPKMIGGIGGFIKVKQYDQILIEICGKKAIGTVLVGPTPVNIIGRNM
                LTLIGCTLNFPISPISTVPVKLKPGMDGPRIKQWPLTEEKIKALTEICADMEREGRISK
                IGPENPYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTV
                LDVGDAYFSVPLDESFRKYTAFTIPSTNNETPGIRYQYNVLPQGWKGSPAIFQASMTKI
                LEPFRSKNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELREHLLKWGFTTPDKKHQKEPP
                FLWMGYELHPDKWTVQPIELPEKESWTVNDIQKLVGKLNWASQIYAGIKVKQLCKLLRG
                TKALTDIVTLTEEAELELAENREILKDPVHGAYYDPSKDLIAEIQKQGQDQWIYQIYQE
                PFKNLKTGKYARKRSAHTNDVKQLAEVVQKVVMESIVIWGKTPKFKLPIQKETWETWWM
                DYWQAT"
CDS             4408..4986
                /gene="vif"
                /codon_start=1
                /product="vif protein"
                /translation="MENRWQVMIVWQVDRMRIRTWNSLVKHHMYISRRAKGWFYRHHYE
                SRHPKVSSEVHIPIGDARIVVRTYWGLQTGEKDWHLGHGVSIEWRLKRYSTQIDPDLAD
                QLIHLHYFNCFSDSAIRKAILGQVVSPRCDYQTGHNKVGSLQYLALKALVTPSRMKPPL
                PSVKKLAEDRWNKPQKTRGRRESHTMNGC"
CDS             4926..5219
                /gene="vpr"
                /codon_start=1
                /product="vpr protein"
                /translation="MEQAPEDQGPQREPYNEWMLDLLEDLKHEAVRHFPRPWLHGLGQH
                IYHTYGDTWEGVEAIIRILQQLLFVHFRIGCQHSRIGINIRGRRVRDGSGRS"
CDS             join(5200..5414,7739..7830)
                /gene="tat"
                /codon_start=1
                /product="tat protein"
                /translation="MDPVDPSLEPWNHPGSQPKTPCNKCYCKVCCYHCQCCFLNKGLGI
                SYGRKKRKPRRGTPQSNKDHQNPIPKQPIPRTQGDSTGPEESKKKVESKTEADRYA"
CDS             join(5339..5414,7739..8035)
                /gene="rev"
                /codon_start=1
                /product="rev protein"
                /translation="MAGRSGNPDEELLRAIRIIKILYQSNPYPEPKGTRQARKNRRRRW
                RARQRQIDTLSERVLSTCLGRPAEPVPLQLPPIERLHLDCSEDCGTSGTQQSQGVETGV
                GRTQVSGESPVVLGSGTKN"
CDS             5441..5686
                /gene="vpu"
                /codon_start=1
                /product="vpu protein"
                /translation="MQLLEICAVVGLVVALIIAIVVWTIVGIEYKKLLKQRKIDRLVDR
                IRERAEDSGNESDGDREELSLLVDMGDYDLGDDNNL"
CDS             5604..8177
```

```
                     /gene="env"
                     /codon_start=1
                     /product="env polyprotein"
                     /translation="MRVMGIERNYPCWWTWGIMILGMIIICNTAENLWVTVYYGVPIWK
                     DANTTLFCASDAKAYDTEVHNVWATHACVPTDPSPQELKMENVTEEFNMWKNNMVEQMH
                     TDIISLWDQSLKPCVQLTPLCVTLDCSYNITNNITNSITNSSVNMREEIKNCSFNMTTE
                     LRDKNRKVYSLFYKLDVVQINNGNNSSNLYRLINCNTSALTQARPKVTFEPIPIHYCAP
                     AGYAILKCNDKEFNGTGLCKNVSTVQCTHGIRPVVSTQLLLNGSLAEGKVMIRSENITN
                     NVKNIIVQLNESVTINCTRPNNNTRRSVRIGPGQTFYATGDIIGDIRQAHCNVSGSQWN
                     KTLHQVVEQLRKYWNNNTIIFNSSSGGDLEITTHSFNCAGEFFYCNTSGLFNSTWVNGT
                     TSSTSNGTITLPCRIKQIINMWQRVGQAMYAPPIQGVIKCESNITGLILTRDGGVNSSD
                     SETFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTKARRRVVEREKRAVTLGAVFIGFL
                     GTAGSTMGAASITLTVQARKLLSGIVQQQSNLLRAIEAQQHLLKLTVWGIKQLQARVLA
                     VERYLRDQQLLGIWGCSGKLICPTNVPWNSSWSNKSLDEIWENMTWLQWDKEISNYTIK
                     IYELIEESQIQQERNEKDLLELDKWASLWNWFDISKWLWYIKIFIMIVGGLIGLRIVFA
                     VLSVINRVRQGYSPLSFQTHTPNPRGLDRPGRIEEEGGEQDRGRSIRLVSGFLALAWDD
                     LRNLCLFSYHRLRDFILIAARTVELPGHSSLKGLRLGWEGLKYLGNLLLYWGRELKISA
                     INLLDTIAIAVAGWTDRVIETVQRLGRAILNIPRRIRQGFERALL"
     CDS             8179..8826
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /translation="MGNKWSKSCIVGWPEVRERIRQTPTAARERTRQAPTAAKGVGAVS
                     QDLDKHGAVTSSNVNHPSCVWLEAQEEEEVGFPVRPQVPLRPMTYKAAFDLGFFLKEKG
                     GLDGLIYSKKRQEILDLWVYHTQGYFPDWQNYTPGPGIRYPLTFGWCFKLVPVDEDEVE
                     EATGGENNSLLHPICQHGMDDEEKETLRWKFDSSLARVHKARELHPEFYKDC"
     misc_signal     8844..8853
                     /note="NFkappaB"
     misc_signal     8858..8867
                     /note="NFkappaB"
BASE COUNT     3238 a   1569 c   2174 g   1999 t
ORIGIN
        1 gacttgaaag cgaaagtaat agggactcga aaacgaaagt tccagagaag aagctctctt
       61 gacgcaggac tcggcttgct gaggtgcaca cagcaagagg cgagagcggc gactggtgag
                                                                  /\major 5' sj
      121 tacgccattt tttgactagc ggaggctaga aggagagaga tgggtgcgag agcgtcagta
                                        gag cds ->
                                    p17 MA start
      181 ttaagtgggg gaaaattaga tgcatgggag aaaattcggt taaggccagg gggaaagaaa
      241 aaatatagat taaaacatct agtatgggca agcagggagc tggaaagatt tgcacttaac
      301 cctagccttt tagaaacaac agaaggatgt caacaaataa tggaacaatt acaatcagct
      361 ctcagaacag gaacagaaga acttagatca ttatataata cagtagcaac cctctattgc
      421 gtacatcaac ggatagaggt aaaagacacc aaggaagctc tagataaaat agaggagata
      481 caaaagaaaa gcaagcaaaa gacacagcag gcagcagctg acacaggaag tagcagcaag
      541 gtcagccaaa attaccctat agtgcaaaat gcacaagggc aaatgatcca ccagtccttg
              gag p17 MA/\gag p24 CA
      601 tcacctagga ctttgaatgc atgggtgaag gtaatagaag aaaaggctct cagcccagaa
      661 gtaataccca tgttctcagc attatcagaa ggagccaccc cacaagattt aaatatgatg
      721 ctgaacatag tgggggggaca ccaggcagct atgcaaatgt taaaagatac catcaatgag
      781 gaagctgcag aatgggacag gctacatcca gtacatgcag ggcctgttgc accaggccag
      841 atgagagaac caaggggaag tgatatagca ggaactacta gtacccctca agaacaaata
      901 gcatggatga caggcaaccc acccatccca gtgggagaca tctataaaag atggatgatc
      961 ctgggattaa ataaaatagt aagaatgtat agccctgtta gcattttgga tataaaacaa
     1021 gggccaaaag aacccttcag agactatgta gataggtttt ttaaaactct cagagctgag
     1081 caagctacac aggaggtaaa aggttggatg acagaaacgt tactgatcca aaatgcaaat
     1141 ccagattgta aatccatcct aagagcatta ggagcagggg ctacattaga agaaatgatg
     1201 acagcatgcc agggagtggg aggacccggc cataaagcaa gagttttggc tgaggcaatg
                                                    gag p24 CA/\gag p2x
     1261 agtcaagtac aacatacaaa cataatgatg cagagaggca attttaaggg ccagaaaagg
                                  gag p2x/\gag p7NC
                                          \gag-pol fusion TF
```

```
     1321 attaagtgtt tcaactgtgg caaagaagga catctagcca aaaattgcag ggctcctaga
     1381 aaaaagggct gttggaaatg tggaagggaa gggcaccaaa tgaaggactg cactgagaga
     1441 caggctaatt ttttagggaa aatctggcct tccagcaaag ggaggccagg aaatttcct
                  -- ----          ^^^^^^^  ^^      ^ ^ ^^^^^^^
gag-pol ribosomal frame shift       stem-loop inverted repeats
                                                          gag p7NC/\p6
     1501 cagagcagac cagaaccaac agccccacca gcagcagaga tctttgggat gagggaagag
     1561 atagtctccc ctccgaagca ggagcagaac gacagggacc agaacccacc ttcagtttcc
                                                            gag-pol TF/\pol
                                                                  protease
     1621 ctcaaatcac tctttggcaa cgacctcttg tcacagtaaa gatagggggа cagctaaaaa
                                           <- gag cds
                                              p6 end
     1681 aagctctatt agatacagga gcagatgata cagtattaga agacataaat ttgccaggaa
     1741 aatggaaacc aaaaatgata gggggaattg gaggtttcat caaggtaaag cagtatgatc
     1801 agatacttat agaaatttgt ggaaaaaagg ctataggtac agtattggta ggacctacac
     1861 ctgtcaacat aattggaaga aatatgttga ccctgattgg ttgtacttta aatttcccaa
                                                      pol protease/\pol RT
     1921 ttagtcctat tagtactgta ccagtaaaat taaaaccagg aatggatggc ccaaggatta
     1981 aacaatggcc attgacagaa gaaaaaataa aagcattaac agaaatttgt gcagatatgg
     2041 aaagagaagg aagaatttca aaaattgggc ctgaaaatcc atacaatact ccaatatttg
     2101 ctataaagaa aaaggacagc actaaatgga gaaaattagt agattttaga gagctcaata
     2161 aaagaactca agacttttgg gaagttcaat taggaatacc gcatccagcg ggcttaaaaa
     2221 agaaaaaatc agtaacagta ctagatgtgg gggacgcata tttttcagtt cccttagatg
     2281 aaagctttag aaagtatact gcattcacca tacctagtac aaacaatgag acaccaggaa
     2341 tcaggtatca gtacaatgtg cttccacagg gatggaaagg atcaccggca atattccagg
     2401 ccagcatgac aaaaatctta gagcccttta gatcaaaaaa tccagacata gttatctatc
     2461 aatacatgga tgacttgtat gtaggatctg atttagaaat agggcagcat agaacaaaaa
     2521 tagaagaatt aagagaacat ctattaaaat ggggatttac tacaccagac aaaaagcatc
     2581 agaaagaacc tccatttctt tggatgggat atgaactcca tcctgataag tggacagtcc
     2641 aaccgataga gctgccagaa aaggaaagct ggactgtcaa tgatatacag aaattagtag
     2701 ggaaactaaa ttgggcaagt caaatttatg caggaattaa agtaaaacaa ttgtgtaaac
     2761 tcctcagggg aaccaaagca ttaacagata tagtaacatt gactgaggaa gcagaattag
     2821 aattggcaga gaacagggag attttaaaag accctgtgca tggagcatat tatgacccat
     2881 caaaagactt aatagcagag atacagaaac aagggcaaga ccaatggata tatcaaattt
     2941 atcaagagcc atttaaaaat ctaaaaacag gaaaatatgc aagaaaaagg tctgctcaca
     3001 ctaatgatgt aaaacaattg gcagaagtgg tgcaaaaggt ggtcatggaa agcatagtaa
     3061 tatggggaaa gactcctaaa tttaaactac ccatacaaaa agaaacatgg gaaacatggt
     3121 ggatggacta ttggcaggct acctgaattc ctgaatggga gtttgtcaat acccctcctc
                                  <- pol cds premature end
     3181 tagtaaaatt atggtaccag ttagagaaag accccatagc aggagcagag actttctatg
                                                          pol RT/\pol
                                                                RNAse
     3241 tagatggggc agccaatagg gagactaagc taggaaaagc agggtatgtc actgacagag
     3301 gaaggcaaaa ggttgtttcc ctaactgaga caacaaatca aaagactgaa ttacatgcaa
     3361 tccatctagc cttgcaggat tcaggatcag aagtaaacat agtaacagac tcacagtatg
     3421 cattaggaat cattcaggcc cagccagaca ggagtgaatc agagttagtc aatcaaataa
     3481 tagagaagct aatagaaaag gacaaagtct acctgtcatg ggtaccagca cacaaaggaa
     3541 ttggaggaaa tgaacaagta gataaattag tcagttctgg aatcaggaag gtactatttt
                                                        pol RNAse/\pol
                                                              integrase
     3601 tagatgggat agataaagct caagaagaac atgaaagata tcacagcaat tggagagcaa
     3661 tggctagtga ttttaatctg ccacctatag tagcaaagga aatagtagcc agctgtgata
     3721 aatgtcagct aaaaggggaa gccatgcatg gacaagtaga ctgtagccca gggatatggc
     3781 aattagattg tacacatcta gaaggaaaag taattctggt agcagtccat gtggctagtg
     3841 gctacgtaga agcagaagtt attccagcag aaacaggaca ggagacagca tactttctac
     3901 taaagctagc aggaagatgg ccagtaaaag tagtacacac agacaatggc agcaatttca
     3961 ccagcgctgc ggttaaagca gcctgttggt gggcaaatgt taaacaggaa tttggtattc
     4021 cctacaatcc ccaaagtcaa ggagtagtag aatctatgaa taaggaatta aagaaaatca
     4081 tagggcaggt aagggagcaa gctgaacatc ttaagacagc agtacaaatg gcagtgttca
     4141 ttcacaattt taaaagaaaa ggggggattg gggggtacag tgcaggggaa agaataatag
```

```
     4201 acataatagc atcagactta caaactaaag aattacaaaa acaaattaca aaaattcaaa
     4261 aatttcgggt ttgttacagg gacagcagag atccaatttg gaaaggacca gcaaaactac
     4321 tctggaaagg tgaaggggca gtggtaatac aggacaatag tgatataaaa gtagtaccaa
     4381 gaagaaaagt aaagatcatt aaggattatg gaaaacagat ggcaggtgat gattgtgtgg
                        vif cds start ->
     4441 caggtagaca ggatgaggat tagaacatgg aacagtctag taaaacatca tatgtatatc
                                <- pol cds normal
                                   integrase end
     4501 tcacggagag ctaaaggttg gttttataga catcactatg aaagcaggca tccaaaagta
     4561 agctcagaag tacacatccc aataggggat gctagaatag tagtaagaac atattggggt
     4621 ctgcagacag gagaaaaaga ctggcacttg ggtcatgggg tctccataga atggaggcta
     4681 aaaagatata gcacacaaat agaccctgac ctggcagacc aactaattca tctgcattat
     4741 tttaactgtt tttcagactc tgccataagg aaagccatat tagggcaagt agttagccct
                  3'sj /\
     4801 aggtgtgatt atcaaacagg acacaacaag gtaggatctt tacaatattt agcactaaaa
                                   5'sj / \
     4861 gcattagtaa caccatcaag gatgaagcca cctttgccta gtgttaagaa attagcagag
     4921 gatagatgga acaagcccca gaagaccagg ggccgcagag agagccatac aatgaatgga
 vpr cds start ->
     4981 tgttagatct gttagaagat cttaagcatg aagctgttag acattttcct aggccatggc
             <- vif cds end
     5041 tccatggatt agggcaacat atctatcaca cctatgggga tacttgggaa ggagttgaag
     5101 ctataataag aattttgcag caactactgt ttgtccattt cagaatcggg tgccaacaca
     5161 gcagaatagg cattaatatt cgagggagaa gagtcaggga tggatccggt agatcctagc
                                     tat cds start - >                  <-
                                                                  vpr cds end
     5221 ctagagccct ggaaccatcc gggaagtcag cctaaaactc cttgtaacaa gtgttactgt
     5281 aaagtgtgtt gctatcattg ccaatgctgc tttctgaaca agggcttagg catctcctat
                                                          rev cds start ->
     5341 ggcaggaaga agcggaaacc ccgacgagga actcctcaga gcaataagga tcatcaaaat
     5401 cctataccaa agcagtaagt atcagtaatt agtatatgta atgcagcttt tggaaatctg
tat-rev exon/intron 5'sj/\              vpu cds start ->
     5461 tgcagtagta ggactggtag tagcgctaat catagcaata gttgtgtgga ctatagtagg
     5521 tatagaatat aagaaattgc taaagcaaag aaaaatagac aggttagttg atagaataag
     5581 agaaagagca gaagacagtg gcaatgagag tgatggggat agagaggaat tatccttgct
                          env cds ->
             signal peptide start
     5641 ggtggacatg ggggattatg atcttgggga tgataataat ttgtaatact gcagaaaact
                                       env signal peptide/\env gp120
                                                  <- vpu cds end
     5701 tgtgggttac tgtctactat ggggtaccta tatggaagga tgcaaatacc accttatttt
     5761 gtgcatcaga tgcgaaagca tatgatacag aagtgcataa tgtctgggct acgcatgcct
     5821 gtgtacctac agaccccagc ccacaagaac taaagatgga aaatgtgaca gaagagttta
     5881 acatgtggaa aaataacatg gtagagcaga tgcatacaga tataatcagt ctatgggacc
     5941 aaagcctaaa accatgtgta cagttaaccc ctctctgcgt tactttagat tgtagctata
     6001 acatcaccaa taatatcacc aatagcatca ccaatagctc agttaacatg agagaagaaa
     6061 taaaaaactg ctctttcaat atgaccacag aattaaggga taagaatcgg aaggtatatt
     6121 cacttttta taaacttgat gtagtacaaa ttaataatgg taataacagt agtaatctgt
     6181 atagattaat aaattgtaat acctcagccc ttacacaggc tcgtccaaag gtaacctttg
     6241 agccaattcc catacattat tgtgccccag ctggttatgc gattctaaaa tgtaatgata
     6301 aggagttcaa tggaacaggg ctatgcaaga atgtcagcac agtacaatgc acacatggaa

     6361 tcaggccagt agtgtcaact caactgctgt taaatggcag tttagcagaa ggaaaggtaa
     6421 tgattagatc tgaaaatatc acaaacaatg tcaaaaacat aatagtacaa cttaacgagt
     6481 ctgtaacaat taattgtacc agacctaaca ataatacaag aagaagtgta cgtataggac
           V3 loop start ->
     6541 caggacaaac attctatgca acaggtgata taatagggga tataagacaa gcacattgta
                                                                  <- V3 loop
                                                                     end
     6601 atgtcagtgg gtcacaatgg aataaaactt tacaccaggt agttgaacaa ttaagaaaat
     6661 attggaacaa caatacaata atctttaata gctcctcagg aggggattta gaaattacaa
```

```
6721 cacatagttt taattgtgca ggagaatttt tctattgtaa tacatcaggc ctgtttaata
6781 gtacttgggt aaatggcact accagcagca cgtcaaatgg cactataact ctcccatgca
6841 gaataaagca aattataaat atgtggcaga gagtaggaca agcaatgtat gcccctccca
6901 tccaaggagt aataaagtgt gaatcaaaca ttacaggact aatattaaca agagatggtg
6961 gggttaatag cagtgacagt gaaaccttca gacctggagg aggagatatg agggataatt
7021 ggagaagtga attatataag tataaggtag taaaaattga accactagga gtagcaccca
7081 ccaaggcaag gagaagagtg gtggagagag aaaaaagagc agttacactg ggagctgtat
                                    env gp120/\env gp41
7141 tcattgggtt cttaggaaca gctggaagca caatgggcgc ggcgtcaata acgctgacgg
7201 tacaggccag aaaattattg tctggcatag tgcaacagca aagcaatttg ctgagggcta
7261 tagaggctca acagcatctg ttgaaactca ctgtctgggg cattaaacag ctccaggcaa
7321 gagtcctggc tgtggaaaga tacctaaggg atcaacagct cctaggaatt tggggctgct
7381 ctggaaaact catctgcccc actaatgtgc cctggaactc tagttggagt aataaatctc
7441 tagatgaaat atgggaaaac atgacctggc tgcaatggga taaagaaatt agcaattaca
7501 caatcaaaat atatgagcta attgaagaat cgcagatcca gcaggaaagg aatgaaaaag
7561 acttactgga gttggacaag tgggcaagtc tgtggaattg gtttgacata tcaaaatggc
7621 tgtggtatat aaaaatattt ataatgatag taggaggcct aataggatta agaatagttt
7681 ttgctgtgct ttctgtaata aatagagtta ggcagggata ctcaccccta tcgtttcaga
                                    tat and rev cds resume (exon 2) /\
7741 cccatacccc gaacccaagg ggactcgaca ggcccggaag aatcgaagaa gaaggtggag
7801 agcaagacag aggcagatcg atacgcttag tgagcgggtt cttagcactt gcctgggacg
                                  <- tat cds end
7861 acctgcggaa cctgtgcctc ttcagctacc accgattgag agacttcatc ttgattgcag
7921 cgaggactgt ggaacttccg ggacacagca gtctcaaggg gttgagactg gggtgggaag
7981 gactcaagta tctggggaat ctcctgttgt attggggtcg ggaactaaaa attagtgcta
                                                                <- rev cds end
8041 ttaatttgct tgataccata gcaatagcag tagctggctg gacagatagg gttatagaaa
8101 cagtacaaag gcttggtaga gctattctca acatacctag aagaatcagg cagggcttcg
8161 aaagggcttt actataacat gggtaacaag tggtcaaaga gttgcatagt gggatggcct
                   <- env cds
                      gp41 end
         nef cds start ->
8221 gaggttaggg aaagaataag acaaactcct acagcagcaa gggaaagaac aagacaagcc
          ---- ---------- ---------- ---------- ---------- -------
                 direct repeat copy 1           direct repeat copy 2
8281 cctacagcag caaaaggagt aggagcagta tctcaagatt tagataaaca tggagcagtc
8341 acaagcagca atgtaaatca ccctagttgc gtctggctgg aagcgcagga ggaagaagag
8401 gtaggctttc cagtcagacc acaggtacct ctaaggccaa tgacttacaa ggcagctttc
8461 gatctcggct tcttttaaa agaaaagggg ggactggatg ggttaattta ctccaagaaa
                  3' LTR U3 region start ->
8521 agacaagaaa tccttgatct gtgggtctac cacacacaag gctacttccc tgattggcag
8581 aattacacac cagggccagg gatcagatac ccactaacat ttggatggtg cttcaagcta
8641 gtaccagtgg atgaagatga agtagaggaa gctactggag gagagaacaa tagcctatta
8701 caccctatat gccaacatgg aatggatgat gaggagaaag aaacattaag gtggaagttt
8761 gacagcagcc tggcacgagt acacaaagca agagagctgc atccggagtt ttacaaagac
8821 tgctgacaca ggagttgctg actgggactt tccgctgggg actttccagg ggaggtgtgg
        <- nef cds end       ------- ---   ------------
                                  NF kappa B sites
8881 tttgggcgga gttggggagt ggctaaccct cagatgctgc atataagcag ctgcttctcg
8941 cttgtactgg gtctctcttg gtagaccaga tcgagcctgg
LTR U3 region/\R rpt
```

LOCUS	COMMON	ACCESSION	LENGTH	REGION	FIRST AUTHOR	REFERENCE

HIVU53175 HIVCA9 U53175 422 bp gag Braaten, D. JVI 70, 4220 (1996)
Comment: This study compared the five lineages of primate immunodeficiency viruses and found that only HIV-1 group M requires cyclophilin A for replication. HIV-1 gag binds to cyclophilin A and incorporates it into virions. If this process is disrupted, virion infectivity is inhibited. Cloned isolates from clades A, B, and D of HIV-1 group M and two clones from group O were used. This first phenotypic difference between group M and group O is consistent with the idea that the two HIV-1 groups were introduced to humans by different zoonotic transmissions. This sequence was obtained by PCR from human PBMCs infected with HIVCA9, a group O virus from Cameroon. GenBank accession number U53175.

HIVU54771 CM240 U54771 9203 bp comp. gen. Carr, J.K. JVI 70, 5935 (1996)
Comment: Blood from an asymptomatic heterosexual 21-year-old Thai man was transported from Thailand to the USA where PBMCs were separated and co-cultivated with PHA-stimulated donor PBMCs. DNA from p24 antigen-positive culture was used to amplify the proviral DNA. The complete genomic sequence of the provirus was determined by the compilation of three clones containing different parts of the viral genome. CM240 is an example of a Thai subtype E virus, which is a mosaic of a clade A virus and a clade E virus. GenBank accession number U54771.

SIU17646 SIVsmSL92a U17646 825 bp gag Chen, Z. JVI 70, 3617 (1996)
Comment: Eight SIV gag and env sequences from sooty mangabeys. Six of them represent new viruses from West Africa. All belonged to the SIVsm/HIV-2 family. The characterization of these sequences supports the hypothesis that each HIV-2 subtype in West Africans originated from widely divergent SIVsm strains, transmitted by independent cross-species events in the same geographic locations. GenBank accession numbers U17646,U48810-U48824.

HIVU52194 U52194 325 bp gag Diaz, R.S. Unpublished (1996)
Comment: This work also includes ten envelope sequences. GenBank accession numbers U52194 and U51942,U52055-U52603.

HIVU56888 1C U56888 402 bp gag Fultz, P.N. Unpublished (1996)
Comment: A set of 12 gag sequences and 12 env sequences from a study on the diversification of two HIV-1 LAI Subtype B strains during the dual infection of a chimpanzee for nine years. GenBank accession numbers U56866-U56899.

HIVU51188 90CR402 U51188 9843 bp comp. gen. Gao, F. JVI 70, 7013 (1996)
Comment: One of a set of three complete genomes from a study linking the HIV-1 epidemic in the heterosexual population in Thailand to an A/E recombinant. 90CR402, previously named CAR-E 4002, was obtained from a man from Bangui, Central African Republic, who had lymphadeopathy, diarrhea, severe weight loss and recurrent respiratory infections. He was infected through heterosexual contact, but the year of infection is unknown. 90CR402 (U51188) was first adapted to growth in chimpanze cells, expanded in chimpanzee cells, and then re-expanded in human PBMCs before lamba cloning and sequencing. 93TH253 (U51189) is from a 21-year-old man from Chiang Mai, Thailand and was previously named CMU010 or 302053. The patient had end-stage AIDS. The mode and year of infection are unknown. 93TH253 was isolated and expanded in human PBMCs, then expanded in H9 cells followed by lambda cloning and sequencing. 92UG037 (U51190) was obtained from WHO-NIAIDS and comes from an asymptomatic 31-year-old female from Entebbe, Uganda, early in infection. The year of infection is unknown and the mode of infection was heterosexual contact. 92UG037 was established and propagated by cocultivation with normal donor lymphocytes, and then PCR amplified and sequenced. 93TH253 and 90CR402 are subtype E (A/E recombinant). Comparison of the two strains showed that they were derived from a common A/E recombinant ancestor, presumably from Central Africa. 92UG037 is subtype A. LTR sequences are available under accession numbers U51282-U51297.

SIVCPZANT CPZANT U42720 8182 bp comp. gen. Vanden Haesevelde, M. Virology 221, 346 (1996)
Comment: A simian immunodeficiency virus phylogenetically linked to HIV-1 which was isolated from a captured wild chimpanzee from Zaire. This is the third SIV strain linked to HIV-1 after SIVCPZ-GAB (X52154) and SIVCPZ-GAB2 (U11495) isolated from chimps in Gabon. While SIVCPZ-ANT and SIVCPZ-GAB share a common ancestor with HIV-1, they differ in that the SIVCPCZ strains have highly conserved V3 regions and HIV-1 has a highly variable V3 region. SIVCPZ-ANT is considered to be an outgroup of HIV-1 and is used to suggest the possibility of various introductions of HIV-1 into the human population.

HO74C12DG 074 Z76649 390 bp gag Kampinga, G.A. Unpublished (1996)
Comment: A set of 224 sequences from an extensive HIV-1 mother-to-child-transmission study in Rwanda. The study covers heterogeneity and coinfection with subtypes A and C. GenBank accession numbers Z75969-Z75978, Z75989-Z75997, Z76010-Z76019, Z76030-Z76038, Z76043, Z76045, Z76047, Z76058-Z76069, Z76071-Z76073, Z76084-Z76093, Z76104-Z76113, Z76123-176133, Z76144-Z76152, Z76154-Z76159, Z76162-Z76166, Z76177-Z76182, Z76184-Z76187, Z76208-Z76217, Z76232, Z76246, Z76249-Z76261, Z76274, Z76283, Z76294, Z76302, Z76312-Z76321, Z76343-Z76352, Z76363-Z76372, Z76383-Z76392, Z76413-Z76422, Z76433-Z76442, Z76453-Z76462, Z76649, Z76706, Z76707, Z76733.

HIVU70271 LIT9485 U70271 210 bp gag Liitsola, K. J. Infect. Dis. In press (1996)
Comment: One of thirty-four Baltic and Russian sequences derived from PCR amplified DNA from PBMCs. GenBank accession numbers U70271-U70304.

HIVMCK1 MCK1 D86068 9752 bp comp. gen. Iwatani, Y. Unpublished (1996)
Comment: This sequence is ~98% identical to HXB2 and other IIIB lab strains of HIV-1. PM213 is another complete genome also ~98% identical to HXB2. GenBank accession numbers D86068 and D86069.

HBABMNT BAB-MNT X99948 387 bp gag Narwa, R. JVI 70, 4474 (1996)
Comment: A set of 54 sequences from a mother-to-child-transmission study. The encoding regions of the matrix protein, p17, were sequenced from viral isolates from 22 nontransmitting mothers, 12 HIV-1 positive mother-and-child pairs, 4 infected children, and 4 transmitting mothers. All patients were attending urban hospitals in Paris. Blood samples taken during delivery for mothers and the first month of life for infants. Most of the European mothers were drug users or sexual partners of drug users. The main route of infection for African mothers was heterosexual intercourse. Sequences were classified among the A, B and G subtypes. GenBank accession numbers X99948-X99949, Z79527-Z79578.

HIVETH2220 ETH2220 U46016 9031 bp comp. gen. Salminen, M.O. ARHR 12, 1329 (1996)
Comment: U46016 is the first reported (almost full length) subtype C sequence from Ethiopia. In its genomic organization, this clone closely resembles subtype A, B, and D isolates except that the core promoter contains three potential binding sites for the transcription factor NF-kB instead of containing two.

HIVKUMM1C KUMM1 L42019 410 bp gag Voevodin, A. ARHR 12, 641 (1996)
Comment: Six sequences (I1-5, KUMM2) from Indian expatriates in Kuwait and one sequence from an Ethiopian expatriate in Kuwait used in a study of the diversity of subtypes found in India, Ethiopia, and Kuwait using gag nucleotide sequences. Voevodin et al. found that isolates KUMM1, KUMM2, I1, I3, I4, and I5 clustered with subtype C while the rest clustered with subtype B sequences. GenBank accession numbers U42013, L42014, L42016-L42019, L42022.

LOCUS	COMMON	ACCESSION	LENGTH	REGION	FIRST AUTHOR	REFERENCE
HIVU53606	MB.01	U53606	306 bp	pol (protease)	Barrie, K.A.	Virology 219, 407 (1996)

Comment: This study analyzed HIV-1 gag and pol sequences from PBMCs from HIV-1-infected mothers and their children. Sixty protease alleles from 12 individuals differed by from 3 to as many as 10 amino acids. Protease variants with a proline at position 63, a substitution associated with resistance to protease inhibitors, appeared in the absence of antiprotease therapy in 7 patients and were transmitted by 2 mothers to their infants. Gag p7 p6 regions were more variable than protease. GenBank accession numbers for the set of 61 sequences are U53606-U53666.

HIVU54771	CM240	U54771	9203 bp	comp. gen.	Carr, J.K.	JVI 70, 5935 (1996)

Comment: Blood from an asymptomatic heterosexual 21-year-old Thai man was transported from Thailand to the USA where PBMCs were separated and co-cultivated with PHA-stimulated donor PBMCs. DNA from p24 antigen-positive culture was used to amplify the proviral DNA. The complete genomic sequence of the provirus was determined by the compilation of three clones containing different parts of the viral genome. CM240 is an example of a Thai subtype E virus, which is a mosaic of a clade A virus and a clade E virus. GenBank accession number U54771.

HIVU50207	HIVZ321B	U50207	1497 bp	gag, pol	Choi, D.J.	Unpublished (1996)

Comment: The virus HIVZ321B is a later passage of HIVZ321 (GenBank accession number M15896), which was isolated from a 1976 Zairian serum sample. Z321 was grown to industrial scale in a chronically infected T-cell line to manufacture an inactivated, therapeutic HIV-1 immunogen. Z321B was established from this industrial scale stock.

HIV1U45026	74H9E1.2	U45026	117 bp	pol (rt)	Cleland, A.	JAIDS 12, 6 (1996)

Comment: Sequences of the RT domain of HIV1 from ten plasma and PBMCs samples of two hemophiliac patients. Samples were obtained before, during, and after long-term treatment with ZDV. The appearance of resistance-associated substitutions differed in both patients both in order and timing. The sequences were too short to subtype. GenBank accession numbers U45206-U45664.

HIVU51188	90CR402	U51188	9843 bp	comp. gen.	Gao, F.	JVI 70, 7013 (1996)

Comment: One of a set of three complete genomes from a study linking the HIV-1 epidemic in the heterosexual population in Thailand to an A/E recombinant. 90CR402, previously named CAR-E 4002, was obtained from a man from Bangui, Central African Republic, who had lymphadeopathy, diarrhea, severe weight loss and recurrent respiratory infections. He was infected through heterosexual contact, but the year of infection is unknown. 90CR402 (U51188) was first adapted to growth in chimpanze cells, expanded in chimpanzee cells, and then re-expanded in human PBMCs before lamba cloning and sequencing. 93TH253 (U51189) is from a 21-year-old man from Chiang Mai, Thailand and was previously named CMU010 or 302053. The patient had end-stage AIDS. The mode and year of infection are unknown. 93TH253 was isolated and expanded in human PBMCs, then expanded in H9 cells followed by lambda cloning and sequencing. 92UG037 (U51190) was obtained from WHO-NIAIDS and comes from an asymptomatic 31-year-old female from Entebbe, Uganda, early in infection. The year of infection is unknown and the mode of infection was heterosexual contact. 92UG037 was established and propagated by cocultivation with normal donor lymphocytes, and then PCR amplified and sequenced. 93TH253 and 90CR402 are subtype E (A/E recombinant). Comparison of the two strains showed that they were derived from a common A/E recombinant ancestor, presumably from Central Africa. 92UG037 is subtype A. LTR sequences are available under accession numbers U51282-U51297.

SIVCPZANT	CPZANT	U42720	8182 bp	comp. gen.	Vanden Haesevelde, M.	Virology 221, 346 (1996)

Comment: A simian immunodeficiency virus phylogenetically linked to HIV-1 which was isolated from a captured wild chimpanzee from Zaire. This is the third SIV strain linked to HIV-1 after SIVCPZ-GAB (X52154) and SIVCPZ-GAB2 (U11495) isolated from chimps in Gabon. While SIVCPZ-ANT and SIVCPZ-GAB share a common ancestor with HIV-1, they differ in that the SIVCPCZ strains have highly conserved V3 regions and HIV-1 has a highly variable V3 region. SIVCPZ-ANT is considered to be an outgroup of HIV-1 and is used to suggest the possibility of various introductions of HIV-1 into the human population.

HIVMCK1	MCK1	D86068	9752 bp	comp. gen.	Iwatani, Y.	Unpublished (1996)

Comment: This sequence is ~98% identical to HXB2 and other IIIB lab strains of HIV-1. PM213 is another complete genome also ~98% identical to HXB2. GenBank accession numbers D86068 and D86069.

HIV759RT01 V759 U14902 204 bp pol Quinones, M.E. ARHR 12, 1117 (1996)
Comment: This study looked at sequences from 81 patients from USA, Venezuela, and Spain who were divided into two groups: those with and those without AZT antiretroviral therapy. Several patients had blood drawn on sequential dates to observe change in sequence over time. Patients came from a variety of risk groups, ages, and health status, and a table of patient data is presented. Envelope gene sequences are also available from some of the patients. Not all 81 sequences reported in this paper were released to the database. GenBank accession numbers U14786-U14903, U16764-U16779 and U40533-U40552. The paper also reports on a new sequence analysis software tool for calculating an "evolutionary index" which is a measure of the evolutionary distance between sequences.

HIVETH2220 ETH2220 U46016 9031 bp comp. gen. Salminen, M.O. ARHR 12, 1329 (1996)
Comment: U46016 is the first reported (almost full length) subtype C sequence from Ethiopia. In its genomic organization, this clone closely resembles subtype A, B, and D isolates except that the core promoter contains three potential binding sites for the transcription factor NF-kB instead of containing two.

HIVU53869 PFA330AZT0.2p25 U53869 969 bp pol Tachedjian, G. Virology 212, 58 (1996)
Comment: This sequence was isolated by passing wild-type HIV1 strain PD in the presence of increasing concentrations of AZT and foscarnet in MT-2 cells. Following 25 passes this strain replicated in the presence of 330 micromolar foscarnet and 0.2 micromolar AZT. This protein has three nonpolymorphic substitutions, LYs70Arg, Val75Ile and Lys219Arg, which together give reduced susceptibility to PFA and wild type susceptibility to AZT. The two other sequences in this set also have mutations which give them similar susceptibilities to PFA and AZT. GenBank accession numbers U53669, U53870, and U53871.

SIU65787 SIVmac79A6.1 U65787 1317 bp pol (rt) VanRompay, K.K.A. Unpublished (1996)
Comment: This sequence was derived from an isolate from a rhesus macaque infected with uncloned SIVmac251 following prolonged zidovudine treatment.

HIVU68783 CFSR1 U68783 380 bp pol Wong, J.K. Unpublished (1996)
Comment: Set of 314 sequences of four patients (A, B, C, and D) from San Diego, USA. Samples obtained from the lymph node, spleen, brain, and cerebral spinal fluid of the patients. Sequences are likely from subtype B viruses. GenBank accession numbers for patient A, U68783-U68866; patient B, U68867-U68924; patient C, U68925-U69008; patient D, U69009-U69096.

HIV1AA201 AA201 U64106 738 bp pol (rt) Zheng, N.N. ARHR 12, 1731 (1996)
Comment: Sequences of 91 subtype B HIV-1 isolates obtained between 1990 and 1995 from 23 males from the Sydney region of New South Wales, Australia. These patients were symptomatic or had AIDS, were on long-term antiviral therapy (ZDV and ddI) and showed the development of resistance mutations. Proviral DNA was isolated from PBMCs and the RT region amplified by nested PCR. GenBank accession numbers U64106-U64196.

LOCUS	COMMON	ACCESSION	LENGTH	REGION	FIRST AUTHOR	REFERENCE

HIVU54771 CM240 U54771 9203 bp comp. gen. Carr, J.K. JVI 70, 5935 (1996)
Comment: Blood from an asymptomatic heterosexual 21-year-old Thai man was transported from Thailand to the USA where PBMCs were separated and co-cultivated with PHA-stimulated donor PBMCs. DNA from p24 antigen-positive culture was used to amplify the proviral DNA. The complete genomic sequence of the provirus was determined by the compilation of three clones containing different parts of the viral genome. CM240 is an example of a Thai subtype E virus, which is a mosaic of a clade A virus and a clade E virus. GenBank accession number U54771.

HIVU51188 90CR402 U51188 9843 bp comp. gen. Gao, F. JVI 70, 7013 (1996)
Comment: One of a set of three complete genomes from a study linking the HIV-1 epidemic in the heterosexual population in Thailand to an A/E recombinant. 90CR402, previously named CAR-E 4002, was obtained from a man from Bangui, Central African Republic, who had lymphadeopathy, diarrhea, severe weight loss and recurrent respiratory infections. He was infected through heterosexual contact, but the year of infection is unknown. 90CR402 (U51188) was first adapted to growth in chimpanze cells, expanded in chimpanzee cells, and then re-expanded in human PBMCs before lamba cloning and sequencing. 93TH253 (U51189) is from a 21-year-old man from Chiang Mai, Thailand and was previously named CMU010 or 302053. The patient had end-stage AIDS. The mode and year of infection are unknown. 93TH253 was isolated and expanded in human PBMCs, then expanded in H9 cells followed by lambda cloning and sequencing. 92UG037 (U51190) was obtained from WHO-NIAIDS and comes from an asymptomatic 31-year-old female from Entebbe, Uganda, early in infection. The year of infection is unknown and the mode of infection was heterosexual contact. 92UG037 was established and propagated by cocultivation with normal donor lymphocytes, and then PCR amplified and sequenced. 93TH253 and 90CR402 are subtype E (A/E recombinant). Comparison of the two strains showed that they were derived from a common A/E recombinant ancestor, presumably from Central Africa. 92UG037 is subtype A. LTR sequences are available under accession numbers U51282-U51297.

SIVCPZANT CPZANT U42720 8182 bp comp. gen. Vanden Haesevelde, M. Virology 221, 346 (1996)
Comment: A simian immunodeficiency virus phylogenetically linked to HIV-1 which was isolated from a captured wild chimpanzee from Zaire. This is the third SIV strain linked to HIV-1 after SIVCPZ-GAB (X52154) and SIVCPZ-GAB2 (U11495) isolated from chimps in Gabon. While SIVCPZ-ANT and SIVCPZ-GAB share a common ancestor with HIV-1, they differ in that the SIVCPCZ strains have highly conserved V3 regions and HIV-1 has a highly variable V3 region. SIVCPZ-ANT is considered to be an outgroup of HIV-1 and is used to suggest the possibility of various introductions of HIV-1 into the human population.

HIVU63314 HIV-1MN U63314 275 bp rev Hua, J. Virology In press (1996)
Comment: A sequence of HIV-1 MN env and rev genes from a study on the changes of function resulting from a natural sequence variation in the activation domain of HIV-1 rev. GenBank accession number U63314.

HIVMCK1 MCK1 D86068 9752 bp comp. gen. Iwatani, Y. Unpublished (1996)
Comment: This sequence is ~98% identical to HXB2 and other IIIB lab strains of HIV-1. PM213 is another complete genome also ~98% identical to HXB2. GenBank accession numbers D86068 and D86069.

HIVU57217 T01 U57217 216 bp tat Lorenzo, E. Unpublished (1996)
Comment: A set of 113 env sequences and 88 tat sequences from US patients derived during a study on the variability of tat and env genes in HIV-1. Sequences are subtype B. GenBank accession numbers U57217–U57304 and U57104–U57216 for the env sequences.

HIVETH2220 ETH2220 U46016 9031 bp comp. gen. Salminen, M.O. ARHR 12, 1329 (1996)
Comment: U46016 is the first reported (almost full length) subtype C sequence from Ethiopia. In its genomic organization, this clone closely resembles subtype A, B, and D isolates except that the core promoter contains three potential binding sites for the transcription factor NF-kB instead of containing two.

LOCUS	COMMON	ACCESSION	LENGTH	REGION	FIRST AUTHOR	REFERENCE

HIVCI17D1 CI17D1 U59559 326 bp env Audoly, G. ARHR In press (1996)
Comment: These sequences are from six different AIDS patients suffering from pulmonary tuberculosis at the Pneumology Hospital of Cocody, Abidjan, Ivory Coast. In this study sequences from viral RNA cocultured on donor PBMCs, proviral DNA sequences from each patients' PBMCs after coculture with donor PBMCs, and proviral DNA sequences directly from uncultured PBMCs were determined. For the cocultured samples viral RNA was harvested from the culture supernatant. PCR or RT-PCR was used to amplify the env V3 region, and 4-7 cloned PCR products were sequenced. A total of 66 sequences from the six patients were published. All 66 were subtype A, and intrapatient sequences were more similar than interpatient sequences. GenBank accession numbers U59559–U59624.

HIVU46206 BHGM5 U46206 290 bp env (C2V3) Barbosa, E.F. Unpublished (1996)
Comment: Set of three Brazilian sequences. Sequence U46206 was obtained by PCR amplification from DNA derived from PBMCs of a female blood donor of unknown health status. This sample was classified as subtype F by the HMA method. Sequence U46210 was obtained from a male individual of unknown health status and was also classified as subtype F by the HMA method. A third sequence, U46122, was classified as subtype B.

HIV1U08355 RU103A U08355 675 bp env (V3-V5) Bobkov, A.F. ARHR 12, 251 (1996)
Comment: One of a set of sequences derived from proviral DNA extracted from PBMCs of a Russian patient. V3-V5 region of env gene extended by nested PCR amplification. The patient belonged to a group of 22 seropositive patients infected nosocomially by HIV-1 originating from a single source. RU103A has a V3 sequence identical to 12 other patients in the group. GenBank accession numbers U08355-U08368.

HIV1BUK3A BUK3A U33095 1089 bp env (V1-V5) Bobkov, A.F. ARHR 12, 1385 (1996)
Comment: Sequences from a study on the genetic variability of HIV-1 in the former Soviet Union. U33095 and U33096 come from a blood sample taken November 1992 from a 39-year-old male Caucasian heterosexual (patient BUK) who lived in Uzbekistan. He reported living in Mozambique in 1984-85 where he was admitted to the hospital several times; however, he reported no sexual relationships with the African residents. Proviral DNA extracted from PBMCs; env gene extended by nested PCR amplification. BUK3A (U33095)and BUK4A (U33096) belong to subtype G, but diverge from the subtype G viruses previously reported from Russia. GenBank accession numbers U33095-U33096.

HIV1MLY10A MLY10A U33104 1218 bp env (V1-V5) Bobkov, A.F. ARHR 12, 1687 (1996)
Comment: One hundred and thirty subjects living in the Russian Federation were enrolled in this study between 1992 to 1996. Proviral DNA was extracted from PBMCs and the env gene was extended by nested PCR; amplified products were then cloned. The study showed that subtypes B and G are predominant in the Russian Federation, although subtypes A, C, D, and H are also present. In this article special attention is given to two epidemiologically linked patients SHL (U33106-U33108) and MLY (U33104-U33105) whose genetic subtypes could not be identified clearly by the HMA method. SHL became infected through sexual contact with an HIV-1 positive student from Zaire. MLY became infected by a blood transfusion from SHL after a cesarean section in 1987. MLY child is believed to have acquired the disease through breast feeding. The authors concluded that MLY and SHL belonged to the H subtype after phylogenetic tree analysis.

H92MY14093 92MY14093 U65538 345 bp env (V3) Brown, T.M. ARHR 12, 1655 (1996)
Comment: Blood samples were collected between 1992 and 1993 from 13 IV drug-using prisoners in Kuala Lumpur, Malaysia and one HIV-1 infected baby born in Thailand but adopted by Malaysian parents. PCR products amplified from uncultured PBMCs were directly sequenced. Eleven sequences clustered with the B subtype (U65538-U65548), two grouped in the C subtype (U65549-U65550) and the baby's sequence (U65551) was determined to be subtype E. Subtype C patients were Malaysian nationals who had visited India in 1989 and 1990 for medical treatment. GenBank accession numbers U65538-U65551.

HIV194UG003 94UG003 U44878 275 bp env (C2-V3) Buonaguro, L. JVI 69, 7971 (1995)
Comment: These 10 sequences are from Gulu, northern Uganda. They are direct sequences of PCR products amplified from uncultured PBMCs. Blood samples were drawn in March 1994 from 217 pregnant women attending a clinic in Gulu, northern Uganda. Ages ranged from 17 to 37 years. The 29 seropositive women (13.4% of the 217 tested) were all asymptomatic. Eight sequences were subtype A and two were D (U44881 U44884). Genbank accession numbers U44878–U44887.

HIV1U37030 AR06 U37030 331 bp env (V3) Campodonico, M. ARHR 12, 79 (1996)
Comment: Fourteen sequences from a study examining HIV-1 strains in Rosario, Agentina. The cohort included individuals infected through heterosexual and homosexual contact, blood transfusion, and intravenous drug use. The patients were enrolled in the study in April 1995. Two (U37032, U37033) clustered with subtype F sequences, ten (U37030, U37031, U37034-U37042) with subtype B, and one (U37043) appeared to be a B/F hybrid.

HIVU54771 CM240 U54771 9203 bp comp. gen. Carr, J.K. JVI 70, 5935 (1996)
Comment: Blood from an asymptomatic heterosexual 21-year-old Thai man was transported from Thailand to the USA where PBMCs were separated and co-cultivated with PHA-stimulated donor PBMCs. DNA from p24 antigen-positive culture was used to amplify the proviral DNA. The complete genomic sequence of the provirus was determined by the compilation of three clones containing different parts of the viral genome. CM240 is an example of a Thai subtype E virus, which is a mosaic of a clade A virus and a clade E virus. GenBank accession number U54771.

HIVU53278 CMCH1 U53278 287 bp env Cassol, S. ARHR 12, 1435 (1996)
Comment: A set of sequences from a study on the detection of HIV-1 subtypes A,B,C, and E in India, Myammar, Thailand, China, and Indonesia involving dried blood spots collected in 1992. Envelope gene extended by PCR amplification from DNA of dried blood spots. The patients belonged to a wide variety of risk categories. GenBank accession numbers UU53286, U53291 (subtype A); U53304-U53308, U53310, U53311, U53314-U53317(subtype B); U53278-U53285, U53287-U53290, U53292-U53303 (subtype C); U53309, U53312, U53313 (subtype E).

HIVTW644 TW64-4 U73045 214 bp env (V3) Chang, K.S.S. ARHR In press (1996).
Comment: These 16 sequences represent healthy HIV-1 carriers or AIDS patients from Taiwan. Three subtype B sequences in this set were greater than 97% identical to the HXB2/LAI lab strain of HIV-1 (TW83, U73049; TW271, U73059; and TW335, U73061). The manuscript reports that 123 of 143 sequences from Taiwan were subtype B, but only 27 of the 143 sequences were submitted to the sequence databases. Other subtypes found in Taiwan in this study were E (U73060, U73062 and U73070), C (U73055), F (U67765) and G (U73058). GenBank accession numbers for B subtype are U73045–U73054, U73056, U73057, U73059, U73061 and U73063–U73069.

HIVU50207 Z321B U50207 1497 bp env Choi, D.J. ARHR In press (1996)
Comment: The virus HIVZ321B is a later passage of HIVZ321, which was isolated from a 1976 Zairian serum sample (GenBank accession number M15896). Z321 was grown to industrial scale in a chronically infected T-cell line to manufacture an inactivated, therapeutic HIV-1 immunogen. Z321B contains a mutation in the termination codon of the tat gene and extends further downstream by 48 nucleotides. The authors present evidence for Z321 being an A/G recombinant. GenBank accession numbers U50207 and U50208.

HIV1GP41 BSMADGP41 X83215 372 bp gp41 Cohen, H.J.M. Lancet 345, 856 (1996)
Comment: One of two partial env group O sequences, one of gp41 and the other of gp120. GenBank accession numbers X83215 and X83216.

HIV1U56263 2BD20 U56263 656 bp env (C2-V5) Contag, C.H. Unpublished (1996)
Comment: These sequences come from a study of mother to child transmission of HIV1 in Sweden involving five subtypes. Subtype A: mother 4, and infant 4 (U56274-U56283, U56328); Subtype B: mother 2, infant 2, mother 6, and infant 6 (U56263-U56273, U56288-U56291, U56299, U56300, U56303, U56317-U56321, U56335); Subtype C, mother 5, infant 5, mother 7, and infant 7; Subtype D, mother 3, infant 3 (U56304-U56306, U56310-U56312, U56330); Subtype E, mother 1, infant 1 (U56309, U56329). Proviral DNA derived from PBMCs or plasma.

HIVG134 G134 X90912 873 bp env (V3-V5) Delaporte, E. AIDS 10, 903 (1996)
Comment: In this study, 17 partial HIV-1 env (V3-V5) gene sequences were determined from 17 different patients selected from a study involving serological screening of 7,082 patients in Libreville, Gabon, and 771 pregnant women and 886 asymptomatic adults in Franceville, Gabon. The study took place from 1986-1994 during which the prevalence of HIV infection in Gabon was low (0.7 - 1.6% in pregnant women, 2,1 - 2.6% in the general population) and remained stable. The 17 isolates sequenced represented five different subtypes (A, C, D, F, and G) and Group O. The high sequence diversity, and low prevalence rates are similar to Cameroon, but different from the rest of Africa. Of the 17 strains, 10 were isolated from Libreville. Of these, 4 were from asymptomatic seropositive persons in 1988 and 6 were collected from AIDS patients between 1989 and 1993. The remaining 7 were collected between 1988 and 1989 from AIDS patients in Franceville. GenBank accession numbers X90912 (G134), X90913 (LBV10-5), X90914 (LBV2-3), X90915 (G135), X90916 (G98), X90917 (G41), X90918 (LBV23-10), X90919 (G109), X90920 (G141), X90921 (VI1076), X90922 (VI526), X90923 (VI354), X90924 (VI685), X96526 (VI686), U09665 (LBV21-7), L22953 and U09665 (VI525). Isolate G139 is not available from GenBank.

HIVU43035 U43035 235 bp env (V3) Diaz, R. Unpublished (1996)
Comment: Set of 124 closely related sequences presumably belonging to subtype B. All 124 sequences came from patients from the United States who had been infected by a common source. GenBank accession numbers U29433-U29437, U29956-U30054, and U43035-U43054. Gag sequences from 12 clones are also available in GenBank entries U31573-U31584.

HIVU43986 D1-116 U42986 240 bp env (C2-V3-C3) Diaz, R. ARHR 12, 1291 (1996)
Comment: These sequences come from a study on the lack of dual HIV infection in a transfusion recipient who was exposed to two seropositive blood components. The patient was a 54-year-old male with oat cell carcinoma of the lung who was treated with prednisone at the time of index transfusion. He was transfused in November 1984. Samples of sera and leukocytes from both donors and from the dual recipient were analyzed. GenBank accession numbers U43986-U44023.

HIVU51942 U51942 325 bp env (C2V3-C3) Diaz, R. Unpublished (1996)
Comment: This work also includes a gag sequence. GenBank accession numbers U51942, U52055-U52063 and U52194.

HIV1V3NO1 NO1 X92902 252 bp env (V3) Engelstad, M. ARHR 12,1733 (1996)
Comment: Sequences of 40 individuals of Norwegian origin from four major risk groups. Eleven were homosexual or bisexual men, 11 were IVDUs, 4 were heterosexual partners of IVDUs, 6 were hemophiliacs who had received contaminated blood products, and 8 were heterosexual individuals infected through intercourse with persons from Zambia (3), USA (1), Grenada (1), Italy (2), and South Africa (1). The dates of seroconversion were unknown. Thirty-eight of the patients were asymptomatic and two of the hemophiliacs were classified as CDC IV at the time when the samples were obtained. Thirty-six of these sequences are subtype B (X92902–X92912, X92915, X92916, X92919–X92941) and four are subtype C (X92913, X92914, X92917, X92918). GenBank accession numbers X92902–X92941.

HIVU56866 C499env1 U56866 573 bp env Fultz, P.N. Unpublished (1996)
Comment: A set of 12 env sequences from a study on the diversification of two HIV-1 LAI Subtype B strains during the dual infection of a chimpanzee for 9 years. GenBank accession numbers U56866-U56887. Also, see GenBank accession numbers U56888-U56899 for gag sequences from the same study.

HIVU51188 90CR402 U51188 9843 bp comp. gen. Gao, F. JVI 70, 7013 (1996)
Comment: One of a set of three complete genomes from a study linking the HIV-1 epidemic in the heterosexual population in Thailand to an A/E recombinant. 90CR402, previously named CAR-E 4002, was obtained from a man from Bangui, Central African Republic, who had lymphadeopathy, diarrhea, severe weight loss and recurrent respiratory infections. He was infected through heterosexual contact, but the year of infection is unknown. 90CR402 (U51188) was first adapted to growth in chimpanze cells, expanded in chimpanzee cells, and then re-expanded in human PBMCs before lamba cloning and sequencing. 93TH253 (U51189) is from a 21-year-old man from Chiang Mai, Thailand and was previously named CMU010 or 302053. The patient had end-stage AIDS. The mode and year of infection are unknown. 93TH253 was isolated and expanded in human PBMCs, then expanded in H9 cells followed by lambda cloning and sequencing. 92UG037 (U51190) was obtained from WHO-NIAIDS and comes from an asymptomatic 31-year-old female from Entebbe, Uganda, early in infection. The year of infection is unknown and the mode of infection was heterosexual contact. 92UG037 was established and propagated by cocultivation with normal donor lymphocytes, and then PCR amplified and sequenced. 93TH253 and 90CR402 are subtype E (A/E recombinant). Comparison of the two strains showed that they were derived from a common A/E recombinant ancestor, presumably from Central Africa. 92UG037 is subtype A. LTR sequences are available under accession numbers U51282-U51297.

HIVUGO2116 UGO2116 U27426 2928 bp env Gao, F.J. JVI 70, 1651 (1996)
Comment: The complete gp160 coding region of this isolate was sequenced along with those of others collected at major epicenters of the AIDS epidemic during a study by Gao et al. The 35 members of this representative panel include members of all major sequence subtypes of HIV-1 group M (clades A-G) as well as an inter-subtype recombinant (F/B) from an infected individual in Brazil. In this panel, all subtype E and three subtype G viruses initially classified on the basis of partial env sequences were found to cluster in subtype A in the 3' half of their gp41 coding region, suggesting that they are also recombinant. The biological activity of PCR derived envelope genes was examined in a single round virus infectivity assay. This analysis identified 20 clones, including one from each subtype or recombinant, that expressed fully functional envelope glycoproteins. GenBank accession numbers U27443-U27445, U27434, U27426, U27419, U27413, U27408, U27401, U27399, U04900-U04929, U30312, U43386, L34667, U08797, U08794, U09131, U08801, U08441–U08447, U09127, U09126.

HIVU48855 94CU053 U48855 1416 bp env (V3) Gomez, C.E. ARHR 12, 553 (1996)
Comment: DNA sequence of the gp120 cds of a HIV-1 isolate from a bisexual male, most probably infected in Cuba in 1992 by heterosexual contact. The virus was isolated two years after seroconversion by cocultivation of patient PBMCs with seronegative donor PBMCs. Nested PCR was used to amplify the gp120 gene and the C2-V3 region. Five clones from the C2-V3 region were isolated and studied. The amino acid sequence GRGR in the tip of the V3 loop is reported here for the first time. This atypical sequence suggests that this virus can escape from antibodies directed against the V3 region of other HIV-1 isolates more frequently found in America. The sequence clustered with subtype B sequences SF2 and LAI. GenBank accession number U48855.

SIVCPZANT CPZANT U42720 8182 bp comp. gen. Vanden Haesevelde, M. Virology 221, 346 (1996)
Comment: A simian immunodeficiency virus phylogenetically linked to HIV-1 which was isolated from a captured wild chimpanzee from Zaire. This is the third SIV strain linked to HIV-1 after SIVCPZ-GAB (X52154) and SIVCPZ-GAB2 (U11495) isolated from chimps in Gabon. While SIVCPZ-ANT and SIVCPZ-GAB share a common ancestor with HIV-1, they differ in that the SIVCPCZ strains have highly conserved V3 regions and HIV-1 has a highly variable V3 region. SIVCPZ-ANT is considered to be an outgroup of HIV-1 and is used to suggest the possibility of various introductions of HIV-1 into the human population.

HIVBJ1 BJ1 U61854 255 bp env (V3) Heyndrickx, L. ARHR 12, 1495 (1996)
Comment: These 18 sequences are from female prostitutes, born in either Ghana or Togo, who live in Benin. Fifteen are from directly sequenced PCR products, derived via RT-PCR from patient serum RNA. Three (233, 251 and 253) are from cloned PCR products, also by RT-PCR from serum RNA. Another subtype A sequence (366, U61870) is not included here, because it was nearly identical (254 of 255 bases identical) to the SF170 sequence from 1988, and thus it likely represents a lab artifact. GenBank accession numbers U61854–U61869, U61871 and U61873.

HIVU63314 HIV-1MN U63314 275 bp env Hua, J. Virology In press (1996)
Comment: HIV-1 MN env and rev gene sequences from a study on the changes of function resulting from a natural sequence variation in the activation domain of HIV-1 rev. MN is from the USA. GenBank accession number U63314.

HIVU47562 A1-E1 U47562 608 bp env Hutto, C.J. JVI 70, 3589 (1996)
Comment: These sequences come from a study on twins with different disease courses. Envelope gene was extended by nested PCR amplification from proviral DNA derived from patients PBMCs. Heterozygotic twins who had been perinatally infected were observed during their first 2 years of life. Twin A (U47562-U47572), the first born, remained asymptomatic while twin B (U47573-U47588) developed AIDS when he was 6 months old and died when he was 22 months old. The absence of neutralizing antibodies may correlate with disease progression since twin A developed neutralizing antibodies and twin B did not. The samples were determined to be nonsyncytium inducing and subtype B. Also see GenBank accession numbers U47589-U47613 for tat sequences from the same patient.

HIVMCK1 MCK1 D86068 9752 bp comp. gen. Iwatani, Y. Unpublished (1996)
Comment: This sequence is ~98% identical to HXB2 and other IIIB lab strains of HIV-1. PM213 is another complete genome also ~98% identical to HXB2. GenBank accession numbers D86068 and D86069.

HIV1686EN VI686 X96526 2640 bp env Janssens, W. AIDS 10, 903 (1996)
Comment: A group O sequence from a study on the charactarization of HIV-1 group O viruses. This sequence is from a 1992 sample taken from a Gabonese woman with AIDS (Libreville). The Method of DNA isolation was not described. DNA was PCR amplified and cloned. See also GenBank accession numbers X90912-X90924.

HIVV3REO1 1A-2 D78614 105 bp env (V3) Kakizawa, J. ARHR 12, 561 (1996)
Comment: These sequences come from a study of the diversity of the HIV1 env V3 region in saliva. Eight saliva samples were collected from seven patients. Two were asymptomatic carriers (isolates 7 and 8), three had AIDS-related complex (isolates 2, 4, 5, and 6, 5 and 6 being from the same patient), and two were patients with AIDS (isolates 1 and 3). Saliva samples corresponding to isolates 1 through 5 were collected between April and June, and isolates 6 through 8 were collected in January 1995. Five cases were children with hemophilia (isolates 1-4, 6 and 7) and two patients were homosexual adults (isolates 5 and 8). Proviral DNA was extracted by glass powder method. DNA was then PCR amplified and sequences. GenBank accession numbers: Isolate 1, D78614-D78616; Isolate 2, D78617; Isolate 3, D78618-D78621; Isolate 4, D78622-D78623; Isolate 5, D78624-D78626; Isolate 6, D78627-D78630; Isolate 7, D78631-D78633; Isolate 8, D78634-D78637. These sequences belong to subtype B and, of the previously-reported Japanese strains, are more closely related to SF162 than to MN.

HIVH13958 H13958 L07243 330 bp env (V3) Kalish, M. Unpublished (1996)
Comment: Twenty-five Haitian sequences were PCR amplified from PBMCs. The set includes direct sequences of PCR amplification products, consensus sequences of multiple clones of PCR products plus one direct sequence, and single clones of PCR products. Full length env for some of these have been expressed. All sequences clustered with subtype B. GenBank accession numbers L07145–L07161, L07163–L07165, L07167–L07207, L07209–L07239, L07241–L07246, U08441–U08447.

HIV074CDE 074CDE Z75958 276 bp env (V3) Kampinga, G.A. Unpublished (1996)
Comment: Three hundred and twelve sequences of Rwandan women and their offspring, taken from seven infants and three mothers. Mother 566 was apparently infected with subtype A and subtype C HIV-1. All other sequences are subtype A. GenBank accession numbers for child 566 are Z76160–Z76161, Z76167-Z76168, Z76169-Z76176, Z76233-Z76248, Z76262-Z76273 and Z76717-Z76724; mother 730, Z76353-Z76362; child 564, Z76074-Z76083; mother 226, Z76046; child 538, Z76134-Z76143, Z76373-Z76382, Z76393-Z76412; child 074, Z75958; child 082, Z75998-Z76009, Z76650; child 081, Z75959-Z75968; child 618, Z76198-Z76207; mother 439, Z76048-Z76057, Z76064-Z76068 and Z76070.

HIVJGV10 4664 Z68508 276 bp env (V3) Kuiken, C.L. J.Gen. Virol. 77, 783 (1996)
Comment: These subtype B sequences are from a set containing 15 Dutch homosexuals, 19 Dutch intravenous drug users, 2 German homosexuals, 2 German intravenous drug users, 5 Scottish homosexuals and six Scottish intravenous drug users. The sequences were used in a study of HIV-1 Vpr, Vpu, and V3 regions and how they vary between risk groups. GenBank accession numbers Z68687-Z68693 and Z68508-Z68616.

HIVH0001L H0001L Z67875 270 bp env (V3) Kuiken, L. AIDS In press (1996)
Comment: These subtype B sequences come from a study on the limited intra-subject evolution of the envelope V3 region in the Amsterdam population. Taken from a cohort of homosexual men who seroconverted between 1985 and 1989, these sequences are from direct sequencing of PCR products after RT-PCR from serum DNM. GenBank accession numbers Z67875, Z67876, Z68110, Z68109, Z68015-Z68089, Z67885-Z67960.

HIVU58393 1P74E001 U58393 252 bp env (V3) Leigh-Brown, A.J. Unpublished (1996)
Comment: A set of 73 env sequences from a Scottish study on the variability of env and pol genes in HIV-1 LAI during zidovudine therapy. PCR amplified PBMC DNA were sequenced and are most probably sybtype B. GenBank accession numbers U58393–U58465.

HIVU68496 sample 136 p1 U68496 276 bp env (V3) Leitner, T. PNAS 93, 10864 (1996)
Comment: This set of 26 sequences from 13 samples contains sequences from env V3 and gag p17. The set is derived from an exactly known Swedish transmission history and includes serial samples from some patients. The index case (patient 1) was a Swedish male who is believed to have contracted HIV while visiting Haiti in 1980. Six Swedish females (patients p2, p4, p5, p7, p8, and p11) were infected by patient 1. Two males (patients p6 and p10) were then infected by these females, and two HIV-infected children (patients p3 and p9) were born to the women. The phylogeny of this data set spans the time 1980 to 1994. GenBank accession numbers p1, U68496, U68497, U68509; p2, U68498, U68511; p3, U68499, U68500, U68512, U68513; p5, U68501, U68514; p6, U68502, U68515; p7, U68503, U68516; p8, U68504, U68505, U68517, U68518; p9, U68506, U68519; p11, U68507, U68508, U68520, U68521. Sequences are unavailable for p4 and p10.

HIVU69646 RU3 U69646 258 bp env (V3) Leitner, T. ARHR 12, 1595 (1996)
Comment: This set contains 22 env V3 sequences from Russian infected patients. The set contains both epidemiologically linked and unlinked patient sequences. Subtypes A, B, F, and G were found from hetero-, homo-, and bisexual as well as nosocomially infected patients. All sequences are direct population estimates from uncultured patient PBMCs and were characterized on MT-2, CEM, and Jurkat-tat cells. GenBank accession numbers U69646-U69667.

HIV1U56146 14Pb1-16 U56146 629 bp env (V3-V5) Liu, S-L. Unpublished (1996)
Comment: Set of sequences from a study on different patterns of progression to AIDS from the same source of infection. GenBank accession numbers U56146-U56235.

HIVU54646 3BC01V2 U54646 174 bp env (V1/V2) Lockey, T.D. ARHR 12, 1297 (1996)
Comment: These sequences represent 22 clones from 2 descendant HTLV-IIIB cultures. The HTLV-IIIB stocks were propagated on H9 cells. DNA was isolated and PCR used to amplify the envelope gene. The sequences were then compared to each other, and to other previously isolated clones, to determine the degree of heterogeneity. The heterogeneity found poses questions about the validity of comparing results from different labs. GenBank accession numbers U54646-U54689.

HIVU57104 e01 U57104 360 bp env (C2-V3) Lorenzo, E. Unpublished (1996)
Comment: A set of 113 env sequences and 88 tat sequences from U.S. patients derived from a study on the variability of tat and env genes in HIV-1. Sequences are subtype B. GenBank accession numbers U57104–U57216 and U57217–U57304 for the tat sequences.

HRW890388 RR890388 L76870 270 bp env (V3) Lukashov, V.V. ARHR 12, 951 (1996)
Comment: These sequences are from recent immigrants to the Netherlands from various countries where AIDS is endemic. Unconventional country codes were used in the common names. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced. Individuals originating from African countries belonged to subtypes A, C, and D. Individuals originally from Zaire showed subtypes F and G. GenBank accession numbers L76870-L76913 and L86887.

HIV1ENVAA 131871 L76842 270 bp env (V3) Lukashov, V.V. ARHR 12, 1179 (1996)
Comment: This set of sequences comes from a study that links the origin of the AIDS epidemic in Northern Europe to the United States. HIV1 RNA coding for the V3 envelope region was derived from serum samples taken at seroconversion from 31 homosexual men who seroconverted in 1985-1988 and 12 IVDUs who seroconverted in 1986-1990 in Amsterdam, and 11 IVDUs from Baltimore who seroconverted in 1988-1990. In addition, V3 sequences from 8 IVDUs in Scotland and Germany and 10 homosexual men in Scotland, Germany, the United States, and Russia were used to conclude the existence of a link to the US in the origin of the northern European AIDS epidemic. GenBank accession numbers L76842-L76863.

HIVU48274 1018 U48274 644 bp env (C2-V5) McCutchan, F.E. JVI 70, 3331 (1996)
Comment: Four of these sequences are from Uruguayan servicemen who acquired HIV-1 infections while deployed as United Nations peacekeepers in Cambodia in 1993. Two sequences are from samples collected in 1993 in Thailand. One sequence is from a U.S. serviceman who acquired an HIV-1 infection while deployed in Thailand. All of these patients were asymptomatic when samples were collected. Six sequences were from patients hospitalized with AIDS-related illnesses in northern Thailand. Viral DNA was PCR amplified from patient PBMCs. All sequences are HIV-1 clade E. GenBank accession numbers U48272-U48278 and U48264–U48269.

HIVU57788 KI4803 U57788 1610 bp env McKeating, J.A. Virology 220, 450 (1996)
Comment: HXB2 viruses chimeric for gp120 were constructed from a primary isolate. The primary isolate was KI4803 which comes from Sweden and whose subtype is uncertain. Primary and chimeric viruses were compared to ascertain differences in functionality. XHB2 chimeras showed same patterns of cell tropism and cytopathicity. Both primary and chimeric viruses were sensitive to neutralization by sCD4 and had the same gp120 conformation. GenBank accession numbers U57788-U57794.

HAR20016 AR20016 U68522 258 bp env (V3) Marquina, S. ARHR 12, 1651 (1996)
Comment: Blood samples from four unrelated patients in Buenos Aires, Argentina, were collected in 1993. Two patients had AIDS: one was an intravenous drug user, 21280 (U68524), and the other a promiscuous heterosexual, 21281 (U68525). Two samples came from asymptomatic patients, 20021(U68523) and 20016 (U68522). Direct sequencing of PCR products was done from uncultured PBMCs. Two sequences were of subtype F (21280 and 20016) and one of subtype B (21281). The last sequence (20021) was a B/F recombinant.

HESP11158 ESP1-1158 U62618 402 bp env Mas, A. ARHR 12, 1647 (1996)
Comment: These sequences come from blood samples taken from a 35-year-old man from Spain collected in April and September 1995. The V3 region was PCR amplified from uncultured PBMCs, cloned into PGEM-SZF, and an individual clone sequenced. The sequences belong to group O. The April sequence has GenBank accession number U62618 and the September sequence has number U62617.

HIVU66414 UY726 U66414 313 bp env Medina, R.D. ARHR 12, 1491 (1996)
Comment: These nine subtype B sequences were derived from blood samples obtained in March-April 1995 from nine HIV-1 infected patients in Montevideo, Uruguay. Uncultured PBMCs were lysed and viral DNA was amplified by nested PCR. These sequences show a high degree of heterogeneity in the apex motif of V3. GenBank accession numbers U66414-U66422.

HIV1U29209 HCM9 U29209 219 bp env Menu, E. Unpublished (1995).
Comment: This subtype B sequence is from Ho Chi Minh city, from a woman infected by her HIV seropositive sexual partner who was thought to have been infected while traveling in Europe, Genbank accession number U29209. Three other sequences in this Vietnamese study from IV drug users in Ho Chi Minh city and Dong Nai, and a female prostitute in Can Tho were found to be subtype E. Genbank accession numbers U29206-U29209

HIV1HWCL1 HWCL1 U34049 487 bp env (V3) Montpetit, M. ARHR 11, 1421 (1995)
Comment: This sequence is the first published sequence of subtype A HIV-1 in Canada. The patient had moved from Uganda in 1983, and was diagnosed as HIV+ in 1989. Viral RNA was recovered from archived, stored patient serum by binding viral particles to CD4-coated wells of an ELISA plate. After RT-PCR, products were cloned and 10 clones were sequenced. This sequence is from a single clone, L1.

HIVU48719 HIV1VN4 U48719 322 bp env (V3) Nerurkar, V.R. ARHR 12, 841 (1996)
Comment: A genotypic analysis from IDUs and CSWs in Vietnam to determine if concurrent epidemics with different HIV-1 subtypes are occurring among the high risk-behavior groups. Blood samples were collected from two HIV-1 CSWs and three IDUs in southern Vietnam during April and May 1995. VN1 (U45239) and VN2 (U45240) were from healthy 17 and 25 year-old female prostitutes from Can Tho and An Giang. VN3 (U48720) and VN4 (U48719) were a 43 year-old male IV drug user with pruritus and splenomegaly from Nha Thang and a 31 year-old healthy IV drug user male from Nha Thang respectively. Genomic DNA was extracted. The V3 loop region was then amplified through PCR and sequenced. It was determined that the Vietnam HIV-1 strains belong to subtype E and that the patients had been infected with indigenous strains circulating within Vietnam.

HIVU67348 VI1011 U67348 453 bp env Nyambi, P.N. ARHR In press (1996)
Comment: Five HIV-2 sequences derived for use in an examination of the cross-neutralization characteristics between HIV-1 and SIVcpz and those between HIV-1 and HIV-2. Nyambi et al. found that the cross-neutralization between HIV-1 and HIV-2 was less extensive than between the HIV-1 and SIVcpz. In addition they saw that, though their binding capacity did not readily reflect their neutralizing capacity, a majority of HIV-1 and SIVcpz sera bound to the V3 peptides while the HIV-2 sera lacked the neutralizing antibodies to the homologous isolate and did not have any reactivity to the V3 peptide. GenBank U67348-U67352.

HIV1U52247 P1T1A1 U52247 243 bp env (V1-V2) Palmer, C. Virology 220, 436 (1996)
Comment: Variations in the V1V2 region were followed over time in six HIV1 infected individuals. Peripheral blood samples were obtained from six seropositive males (CDC II) attending an STD clinic. The V1V2 sequences were cloned into HXB2 to produce chimeras. The chimeras showed that the V1V2 region of gp120 can determine both tropism and cytopathicity independently from the V3 region. GenBank accession numbers for patient 1, U52247-U52264; patient 2, U52265-U52285; patient 3, U52286-U52303; patient 4, U52304-U52315; patient 5, U52316-U52333; patient 6, U52334-U52339.

HIV1ES106 ES106 U40533 402 bp env (C2-V3) Quinones, M.E. ARHR 12, 955 (1996)
Comment: These subtype B sequences are from 41 patients sampled in Madrid, Spain between 1985 and 1991. Proviral DNA was extracted from uncultured patient PBMCs and the C2V3 region was PCR amplified. The PCR products were directly sequenced. Two of the sequences reported in this set (D22-28 and D22-48) were 99.5% identical to the LAI strain of HIV-1. Three other groups of sequences had members that were greater than 98% identical to each other (R1, R2 and R3; THF13-2, THF12-24; S1, S4). GenBank accession numbers U40533–U40552 and U45286–U45307.

HIVENVB HIV1BB L78831 1546 bp env Ray, S.C. Unpublished (1996)
Comment: These subtype B sequences come from a study on the cytolytic T lymphocyte strain-specific responses that are directed against the HIV-1 env. Virus was isolated from PBMCs. GenBank accession numbers L78831 and L78832.

HIVKZ21 KZ21 U43097 945 bp partial env Reitz, M. Unpublished (1990)
Comment: These subtype A sequences are from Zaire. GenBank accession numbers U43097–U43100.

HIV1U61875 94TZ1574 U61875 342 bp env (C2V3) Robbins, K.E. ARHR 12, 1389 (1996)
Comment: These sequences come from a study on the genetic variability of HIV1 in rural Northwest Tanzania. Sequences were obtained by PCR amplification of patient PBMCs. Seven of the Tanzanians lived in the Mara region and their samples were collected in 1994. The other patient, 87TZ4622, was originally from the Mara region, but had lived in Kigoma for several years. The specimen used was collected in 1987. Sequences 94TZ1627, 87TZ4622, 94TZ1585, and 94TZ1604 are subtype D. Sequences 94TZ1577, 94TZ1576, 94TZ1574, and 94TZ1584 are subytpe A. GenBank accession numbers U61875-U61881 and U65075.

HIV1U31585 SP203 U31585 648 bp env (V3) Sabino, E.C. AIDS 10, 1579 (1996)
Comment: This is a set of 11 sequences from a study on the prevalence of HIV-1 subtypes in Sao Paulo, Brazil. Five sequences are subtype B (U31586, U31587, U31589-U31591), one sequence is subtype C (U31585), and four sequences are subtype F (U31588, U31592-U31595).

HIVETH2220 ETH2220 U46016 9031 bp comp. gen. Salminen, M.O. ARHR 12, 1329 (1996)
Comment: U46016 is the first reported (almost full length) subtype C sequence from Ethiopia. In its genomic organization, this clone closely resembles subtype A, B, and D isolates except that the core promoter contains three potential binding sites for the transcription factor NF-kB instead of containing two.

HIV1G1X11 G1-11 Z50841 105 bp env (V3) Schreiber, M.G. Unpublished (1996)
Comment: Sequences from a study of the loss of antibodies specific for the V3 domain over time in certain HIV-1 variants. Proviral DNA extracted from patients PBMCs. Probably subtype B. GenBank accession numbers Z50841-Z50847.

HIVU45860 1018674A U45860 280 bp env (C2V3) Schwartz, D.H. Unpublished (1996)
Comment: These subtype B sequences are from a study of the envelope glycoprotein C2-V3 region on isolates obtained from PBMCs of four HIV-1 infected patients. GenBank accession numbers for patient 101867 are U45860-U45876, U49624-U49640; patient 10188B, U51311-U51326; patient 10185W, U49600-U49623; patient 10187Y, U49518-U49553.

HIVU45330 U45330 363 bp env (V3) Sheehy, N. J.Gen.Virol 77, 1071 (1996)
Comment: Set of subtype B sequences from an AIDS patient from the United Kingdom who had received treatment with zidovudine for 9 months. Proviral DNA obtained from PBMCs drawn 13 months and 10 months prior to death and from autopsy samples of cardiac blood, lymph node, spleen, bone marrow and brain was PCR amplified and sequenced. A reduction in genetic heterogeneity of the envelope region of viruses present in the proviral blood population occured during treatment coinciding with increased pol heterogeneity. GenBank accession numbers U45330-U45421.

HIV1U29179 UIND1 U29179 1488 bp env Tripathy, S.P. ARHR 12, 1199 (1996)
Comment: Part of a set of nine sequences from New Delhi and Pune, India, eight of which were subtype C (U29179,U29694-U29698, U31362-U31363) and one of which was subtype B (U31364). All 8 subtype C sequences were from heterosexually infected patients. DNA was isolated from cocultured PBMCs after one week of culture. PCR product was cloned and a single clone was sequenced.

SIVmac251 SIU62333 U62333 1114 bp env Trivedi, P. JVI 70, 6875 (1996)
Comment: Fifty-one env sequences from a study of intrarectal transmission of SIV in rhesus macaque monkeys. Trivedi et al. found that a condition now called transient viremia sometimes occurs after mucosal inoculation of monkeys with low doses of SIV as opposed to the more prevalent persistent infection with evident disease progression. Also, after a 1000 infective dose, they found that transiently viremic monkeys were resistant to an intrarectal SIVmac challenge but not to an intravenous SIVmac challenge. GenBank U62333-U62383.

HIVFF1 FF1 Z76463 303 bp env (V3) van der Hoek, L. J. Gen. Virol. In press (1996)
Comment: These subtype B sequences are of 81 clones (patient N) and 105 clones (patient F) from serum, sigmoid tissue and fecal matter from each patient. All sequences from patient N were more similar to other sequences from patient N than to any other sequence in the database. Likewise all sequences from patient F were most similar to other patient F sequences. Both patients are from the Netherlands. GenBank accession numbers Z76463–Z76648.

HIVIU60152 ID4-77c7 U60152 939 bp env Wang, W.K. ARHR 12, 1195 (1996)
Comment: This study concentrates on the change of cysteine residues in region 1 of gp120 HIV-1 Subtype B. Sequential isolates from eight subjects were used though these 11 sequences account for only one of the subjects. Subjects came from a group of 315 homosexual and bisexual men who have been followed at the Fenway Community Health Center in Boston and were enrolled between January 1985 and June 1986. Viruses were isolated from PBMCs of the subjects by co-cultivation with PBMCs of HIV-1 seronegative donors. GenBank accession numbers U60152-U60162.

HIVBJP23A BJP23 D67089 264 bp env (C2V3) Xin, K. Lancet 346, 1372 (1996)
Comment: This pair of sequences is from a study of a dual infection with HIV1 Thai subtype B and E. D67089 is Subtype E and D67090 is Subtype B or a B/E recombinant. GenBank accession numbers D67089 and D67090.

HIV1U32658 NA111 U32658 285 bp env (V3) Zachar, V. ARHR 12, 75 (1996)
Comment: This was a May-June 1992 study of pregnant women from the Pumwani Maternity Hospital in Nairobi, Kenya. Viral RNA was concentrated from patient serum just prior to delivery, and the envelope C2-V3 region was amplified by RT-PCR. The PCR product was cloned and 20 clones from each patient were sequenced. Patients from this study had viral subtypes A (U32658, U33763, U33764, U33766, U33767, and U34905), C (U33762), and D (U33765).

HIVU53192 BTS11 U53192 285 bp env (V3) Zachar, V. ARHR 12, 1069 (1996)
Comment: Genetic analysis of the V3 env region of virus isolates obtained in 1993 from 11 homosexual men living in the inner city of Bratislava. A 32-year-old homosexual man from Prague was also included in the study. Viruses were cultivated in PBMCs. The C2V3 region of HIV-1 LAI was amplified by nested PCR. PCR product was cloned and eight clones from each patient were sequenced. These strains cluster in the B clade together with the majority of North American and western European strains. GenBank accession numbers U53192-U53203.

HIVU50780 AD39PB1 U50780 1047 bp env (V1-V5) Zhu, T. JVI 70, 3098 (1996)
Comment: DNA extracted from PBMCs of an acute seroconvertor male patient, AD39, who was infected with HIV-1 through homosexual transmission by his partner AD38. Sequences spanned by primers PE1 and P2 from the gp120 gene were expanded by PCR amplification. A second round of PCR amplification expanded V1-V2 sequences (using inner primers P1 and P10) and V3 (using inner primers P5 and PV3). AD39PB1 has been characterized as non-syncytium inducing (NSI). GenBank accession numbers U50780-U50815.

LOCUS	COMMON	ACCESSION	LENGTH	REGION	FIRST AUTHOR	REFERENCE
HIVU48901	122-16	U48901	615 bp	nef	Artenstein, A.W.	ARHR 12, 557 (1996)

Comment: Sequences derived from PCR amplified PBMC DNA from nineteen Thai individuals. Only two clones (28-19, 28-2) from one subject clustered with subtype B sequences while the rest clustered with subtype E sequences. The lack of subtype E nef sequences led Artenstein et al. to begin this study. GenBank accession numbers U48897-U48934.

HIVU54771 CM240 U54771 9203 bp comp. gen. Carr, J.K. JVI 70, 5935 (1996)
Comment: Blood from an asymptomatic heterosexual 21-year-old Thai man was transported from Thailand to the USA where PBMCs were separated and co-cultivated with PHA-stimulated donor PBMCs. DNA from p24 antigen-positive culture was used to amplify the proviral DNA. The complete genomic sequence of the provirus was determined by the compilation of three clones containing different parts of the viral genome. CM240 is an example of a Thai subtype E virus, which is a mosaic of a clade A virus and a clade E virus. GenBank accession number U54771.

HIVU51188 90CR402 U51188 9843 bp comp. gen. Gao, F. JVI 70, 7013 (1996)
Comment: One of a set of three complete genomes from a study linking the HIV-1 epidemic in the heterosexual population in Thailand to an A/E recombinant. 90CR402, previously named CAR-E 4002, was obtained from a man from Bangui, Central African Republic, who had lymphadeopathy, diarrhea, severe weight loss and recurrent respiratory infections. He was infected through heterosexual contact, but the year of infection is unknown. 90CR402 (U51188) was first adapted to growth in chimpanze cells, expanded in chimpanzee cells, and then re-expanded in human PBMCs before lamba cloning and sequencing. 93TH253 (U51189) is from a 21-year-old man from Chiang Mai, Thailand and was previously named CMU010 or 302053. The patient had end-stage AIDS. The mode and year of infection are unknown. 93TH253 was isolated and expanded in human PBMCs, then expanded in H9 cells followed by lambda cloning and sequencing. 92UG037 (U51190) was obtained from WHO-NIAIDS and comes from an asymptomatic 31-year-old female from Entebbe, Uganda, early in infection. The year of infection is unknown and the mode of infection was heterosexual contact. 92UG037 was established and propagated by cocultivation with normal donor lymphocytes, and then PCR amplified and sequenced. 93TH253 and 90CR402 are subtype E (A/E recombinant). Comparison of the two strains showed that they were derived from a common A/E recombinant ancestor, presumably from Central Africa. 92UG037 is subtype A. LTR sequences are available under accession numbers U51282-U51297.

SIVCPZANT CPZANT U42720 8182 bp comp. gen. Vanden Haesevelde, M. Virology 221, 346 (1996)
Comment: A simian immunodeficiency virus phylogenetically linked to HIV-1 which was isolated from a captured wild chimpanzee from Zaire. This is the third SIV strain linked to HIV-1 after SIVCPZ-GAB (X52154) and SIVCPZ-GAB2 (U11495) isolated from chimps in Gabon. While SIVCPZ-ANT and SIVCPZ-GAB share a common ancestor with HIV-1, they differ in that the SIVCPCZ strains have highly conserved V3 regions and HIV-1 has a highly variable V3 region. SIVCPZ-ANT is considered to be an outgroup of HIV-1 and is used to suggest the possibility of various introductions of HIV-1 into the human population.

HIVMCK1 MCK1 D86068 9752 bp comp. gen. Iwatani, Y. Unpublished (1996)
Comment: This sequence is ~98% identical to HXB2 and other IIIB lab strains of HIV-1. PM213 is another complete genome also ~98% identical to HXB2. GenBank accession numbers D86068 and D86069.

HIVU61773 And1.dat U61773 618 bp nef Mariani, R. Unpublished (1996)
Comment: Set of 62 sequences from a German study on the frequency of defective nef alleles in an HIV-1 long term survivor. GenBank accession numbers U61773-U61834.

HIVETH2220 ETH2220 U46016 9031 bp comp. gen. Salminen, M.O. ARHR 12, 1329 (1996)
Comment: U46016 is the first reported (almost full length) subtype C sequence from Ethiopia. In its genomic organization, this clone closely resembles subtype A, B, and D isolates except that the core promoter contains three potential binding sites for the transcription factor NF-kB instead of containing two.

II

Amino Acid Alignments

Alignments in this section of the 1996 compendium were generated in one of two ways: On the one hand, a Hidden Markov Method (HMM) was employed as described in Part III of this compendium (Myers and Farmer). In some instances, hundreds of sequences contributed to the alignment and the generation of a consensus-like (model) sequence. The latter is not an ordinary consensus sequence but rather a "most-likely" sequence, with some bias towards HIV-1 sequences. The BLOCKMAKER program (Henikoff and Henikoff, *Meth. Enzymol.* **266**:88, 1996) was utilized to critique the HMM alignments; the blocks found (usually by the Motif option rather than the Gibbs sampler option) are shown by shading. Also, the "Cobbler" sequence inferred from the blocks is appended to each alignment; this sequence serves as another consensus-like sequence. Because the Cobbler sequence was derived from just the twenty or so representative PIVs, it is not so biased towards HIV-1s. On the other hand, new HIV-1 sequences were added to previous HIV-1 alignments that had been created using the PIMA and MASE tools, as in earlier compendiums. The new subtype consensus sequences were then aligned to one another using MASE.

With few exceptions (most notably with HIV-1 Env), only full-length protein sequences have been included in this compilation. Tables giving the basic information about each sequence contained in these alignments precede the amino acid alignments; these tables also appear in Part I.

Ten sequence subtypes have been identified for HIV-1s (see Part I and Part III) and five sequence subtypes exist for HIV-2s. Sequence subtypes have been defined by "cladistic" criteria applied to nucleotide sequences; it remains to be seen whether the amino acid sequence subtypes inferred from that classification are valid, or whether a "phenetic" classification would be preferable.

The reference sequences for the ordinary alignments are mixed case consensus sequences in which upper case letters refer to amino acid residues which are conserved 100% and lower case letters represent amino acid residues conserved in at least 50% of the sequences. The symbol "?" indicates no consensus at a position. Consensus sequences have been generated for each of the defined subtypes (see Parts I and III), and these are presented both with the grand alignments and in alignment to one another. The user should keep in mind that these subtypes have been "cladistically" defined, not "phenetically" defined (the number of phenotypes remains to be discovered).

Within the sequences, "-" is used to indicate residues conserved with respect to the reference sequence and "." represents actual gaps. The symbol "$" indicates a stop codon and the symbol "#" indicates a frameshift or untranslatable situation. Blank spaces within the alignment indicate lack of sequence information over that region. Annotation of the *env* amino acid sequences utilizes "*" for conserved cysteine residues and "^^^" for potential N-linked glycosylation sites.

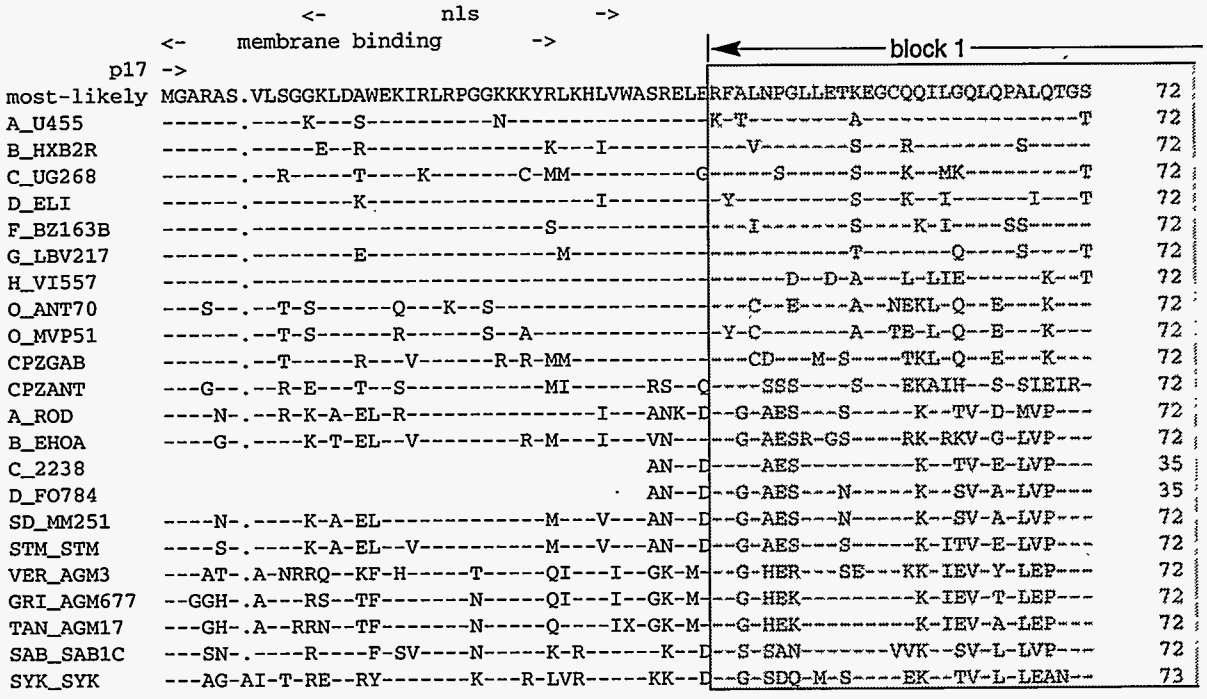
At the risk of relaxing standards of annotation, we have elected to annotate features for which some evidence supports a particular role for a motif—nls for nuclear localization signal, for example. The authority for these additions can be found for the most part in section III curatorial comments (see the Gag entry from 1994 or the Vpr entry from 1995, for example).

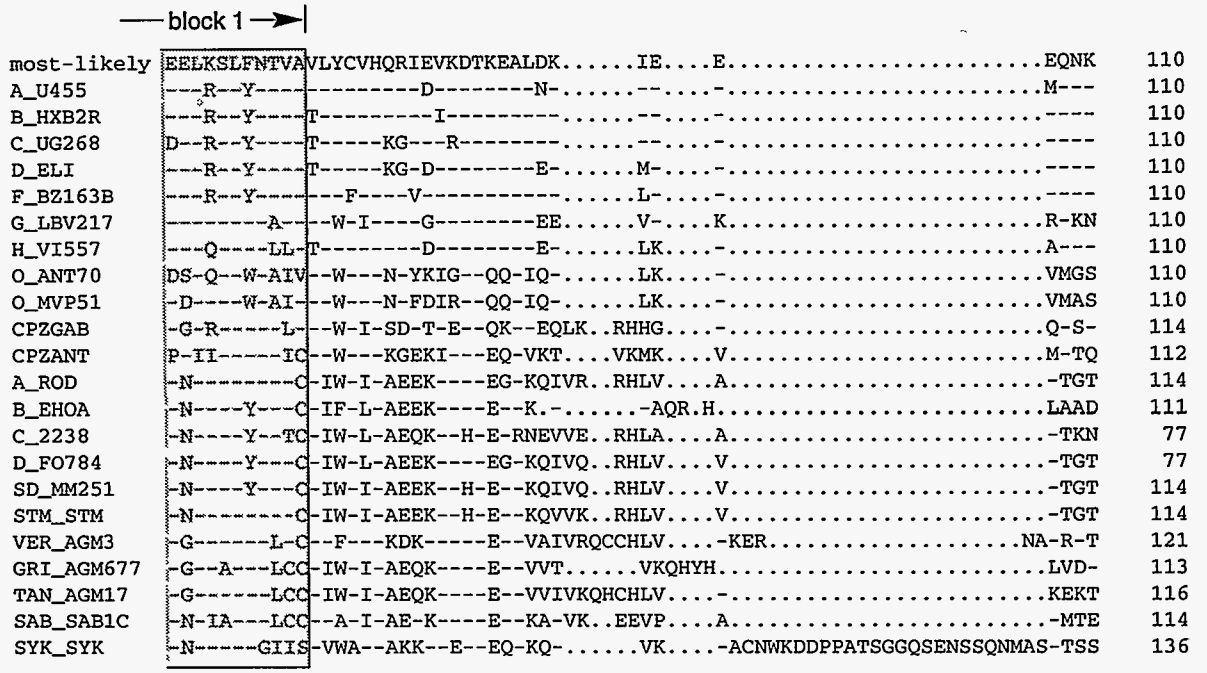
PART II Amino Acid Alignments

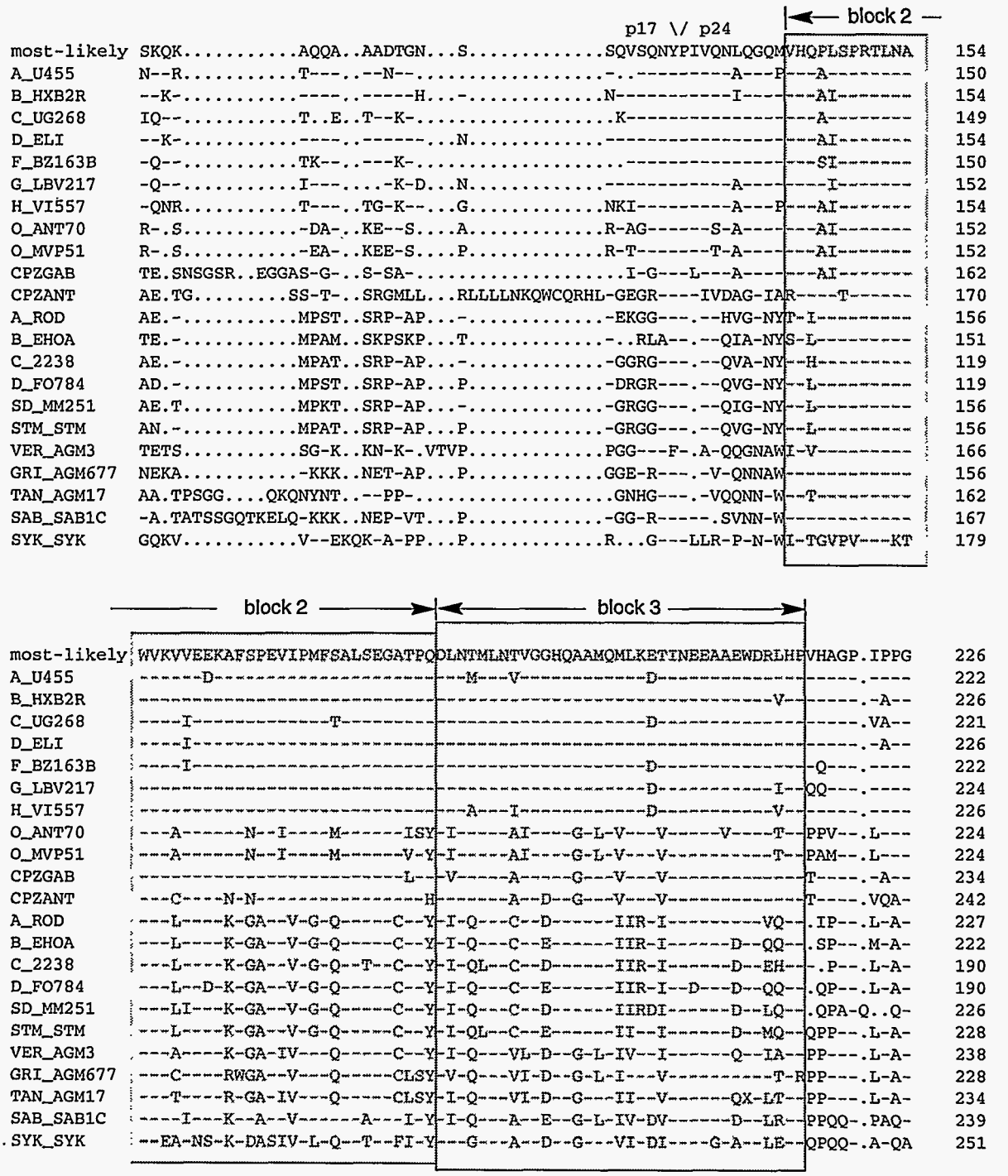
HMMER Sequences in the Gag Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_HXB2R	HIVHXB2R	K03455	Starcick,B.	Science 227, 538 (1985)
C_UG268	HIVUG268	L11799	Louwagie,J.J.	AIDS 7, 769 (1993)
D_ELI	HIVELI	K03454	Alizon,M.	Cell 46, 63 (1986)
F_BZ163B	HIV1BZ163B	S0585	Louwagie,J.J.	ARHR 10, 561 (1994)
G_LBV217	HIVLBV217	L11778	Louwagie,J.J.	AIDS 7, 769 (1993)
H_VI557	HIVVI557	L11793	Louwagie,J.J.	AIDS 7, 769 (1993)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68, 1586 (1994)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
CPZANT	SIVCPZANT	U42720	Vanden Haesevelde,M.	Virology 221, 346 (1996)
A_ROD	HIV2ROD	M15390	Clavel,F.	Nature 324, 691 (1986)
B_EHOA	HIV2EHOA	U27200	Rey-Cuille,M.A.	Virology 202, 471 (1994)
SD_MM251	SIVMM251	M19499	Franchini,G.	Nature 328, 539 (1994)
STM_STM	SIVSTM	M83293	Novembre,F.J.	Virology 186, 783 (1992)
VER_AGM3	SIVAGM3	M30931	Baier,M.	Virology 176, 216 (1990)
GRI_AGM677	SIVAGM677	M66437	Fomsgaard,A.	Virology 182, 397 (1991)
TAN_AGM17	SIVAGM17	L19250	Hirsch	Virology 197, 426 (1993)
SAB_SAB1C	SIVSAB1C	U04005	Jin,M.J.	EMBO J. 13, 2935 (1994)
SYK_SYK	SIVSYK	L06042	Hirsch,V.M.	JVI 67, 1517 (1993)

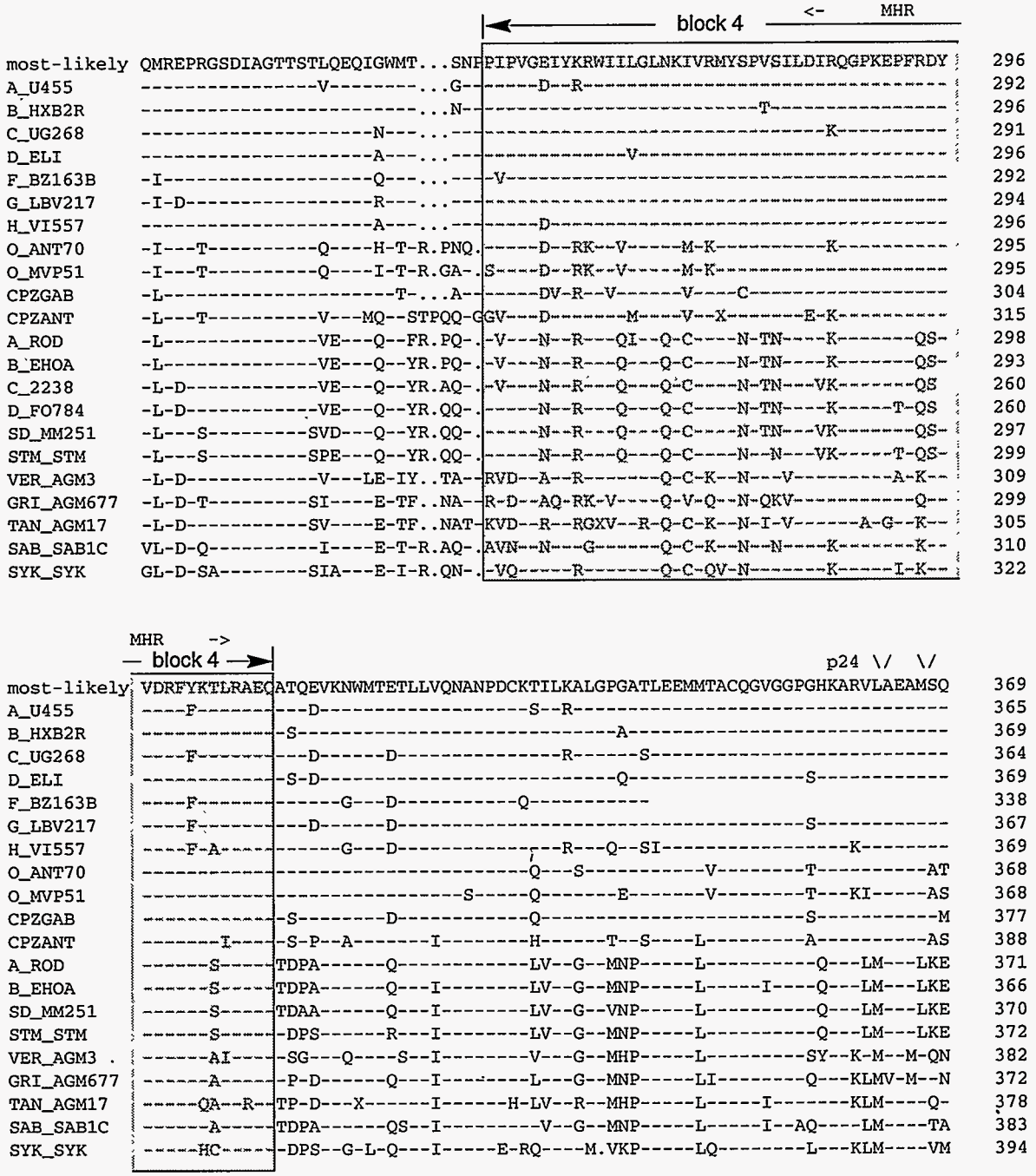
The following alignment and most-likely sequence were generated using the HMMER program as described in Part III. For simplicity, only representative type and subtype sequences are shown. The annotation is based on HIV1s, therefore the user should be cautious about its applicability to other PIV sequences. Nuclear localization signal is indicated by nls. MHR is the major homology region. Cleavage sites are indicated by '\/'. Four conserved, gapless blocks were generated using the BLOCKMAKER and Motif programs; these are shown by shading (The Gibbs program identified five blocks that overlap the four Motif blocks). The Motif "Cobbler" sequence follows this alignment.






p17 \/ p24
block 2
block 2
block 3
most-likely SKQK.....AQQA..AADTGN...S.....SQVSQNYPIVQNLQGQMVHQPLSPRTLNA 154
A_U455 N--R.....T---..--N--.....-..---------A---P---A-------- 150
B_HXB2R --K-.....----..-----H...-.....N----------I-------AI------- 154
C_UG268 IQ--.....T..E..T--K-.....K----------------A-------- 149
D_ELI --K-.....----..------...N.....----------AI------- 154
F_BZ163B -Q--.....TK--..---K-.....----------SI------- 150
G_LBV217 -Q--.....I---....-K-D...N.....----------A---------I------- 152
H_VI557 -QNR.....T---..TG-K--...G.....NKI---------A---P---AI------- 154
O_ANT70 R-.S.....-DA-..KE--S....A.....R-AG------S-A---------AI------- 152
O_MVP51 R-.S.....-EA-..KEE-S....P.....R-T-------T-A---------AI------- 152
CPZGAB TE.SNSGSR..EGGAS-G-..S-SA-.....I-G---L---A---------AI------- 162
CPZANT AE.TG.....SS-T-..SRGMLL...RLLLLNKQWCQRHL-GEGR----IVDAG-IAR----T------ 170
A_ROD AE.-.....MPST..SRP-AP...-.....-EKGG---.--HVG-NYT-I--------- 156
B_EHOA TE.-.....MPAM..SKPSKP...T.....-..RLA--.--QIA-NYS-L--------- 151
C_2238 AE.-.....MPAT..SRP-AP...-.....-GGRG---.--QVA-NY--H--------- 119
D_FO784 AD.-.....MPST..SRP-AP...P.....-DRGR---.--QVG-NY--L--------- 119
SD_MM251 AE.T.....MPKT..SRP-AP...-.....-GRGG---.--QIG-NY--L--------- 156
STM_STM AN.-.....MPAT..SRP-AP...P.....-GRGG---.--QVG-NY--L--------- 156
VER_AGM3 TETS.....SG-K..KN-K-.VTVP.....PGG---F-.A-QQGNAWI-V--------- 166
GRI_AGM677 NEKA.....-KKK..NET-AP...P.....GGE-R---.-V-QNNAW---------- 156
TAN_AGM17 AA.TPSGG....QKQNYNT..--PP-.....GNHG---.-VQQNN-W--T--------- 162
SAB_SAB1C -A.TATSSGQTKELQ-KKK..NEP-VT...P.....-GG-R-----.SVNN-W---------- 167
SYK_SYK GQKV.....V--EKQK-A-PP...P.....R...G---LLR-P-N-WI-TGVPV---KT 179
most-likely WVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRLHPVHAGP.IPPG 226
A_U455 ------D-----M---V----------D---------.---- 222
B_HXB2R ----------V--------.-A-- 226
C_UG268 ----I----------T---------D----------.VA-- 221
D_ELI ----I----------.-A-- 226
F_BZ163B ----I----------D----------Q---.---- 222
G_LBV217 ----------D----------I--QQ---.---- 224
H_VI557 ---A---I----------D----------V--------.---- 226
O_ANT70 ---A------N--I----M------ISY-I-----AI----G-L-V---V-----V----T--PPV--.L--- 224
O_MVP51 ---A------N--I----M------V-Y-I-----AI----G-L-V---V----------T--PAM--.L--- 224
CPZGAB ----------L--V-----A-----G---V---V----------T-----.-A-- 234
CPZANT ---C----N-N----------H-------A--D--G---V---V----------T-----.VQA- 242
A_ROD ---L----K-GA--V-G-Q-----C--Y-I-Q---C--D------IIR-I---------VQ--.IP--.L-A- 227
B_EHOA ---L----K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I------D--QQ--.SP--.M-A- 222
C_2238 ---L----K-GA--V-G-Q--T--C--Y-I-QL--C--D------IIR-I------D--EH--.P--.L-A- 190
D_FO784 ---L--D-K-GA--V-G-Q-----C--Y-I-Q---C--E------IIR-I--D---D--QQ--.QP--.L-A- 190
SD_MM251 ---LI---K-GA--V-G-Q-----C--Y-I-Q---C--D------IIRDI------D--LQ--.QPA-Q..Q- 226
STM_STM ---L----K-GA--V-G-Q-----C--Y-I-QL--C--E------II--I------D--MQ--QPP--.L-A- 228
VER_AGM3 ---A----K-GA-IV---Q-----C--Y-I-Q---VL-D--G-L-IV--I------Q--IA--PP---.L-A- 238
GRI_AGM677 ---C----RWGA--V---Q-----CLSY-V-Q---VI-D--G-L-I---V---------T-RPP---.L-A- 228
TAN_AGM17 ---T----R-GA-IV---Q-----CLSY-I-Q---VI-D--G---II--V------QX-LT--PP---.L-A- 234
SAB_SAB1C ----I---K--A--V------A---I-Y-I-Q---A--E--G-L-IV-DV------D--LR--PPQQ-.PAQ- 239
SYK_SYK --EA-NS-K-DASIV-L-Q--T--FI-Y---G---A--D--G---VI-DI----G-A--LE--QPQQ-.A-QA 251

HIV1 GAG



```
                'p2'        \/ p7                    <-  Zn motif ->      <- Zn motif
most-likely V.TN......ANIMMQ......RG.N.FRGQRKT....IKCFNCGKEGHIARNCRAPRKKGCWKCGKEGHQMK    423
A_U455      -.QQ......TS----.......--.-.---P-R.....----------L-K----------------------    418
B_HXB2R     -.--S.....-T----.......--.-.--N---I....V----------T----------------------    424
C_UG268     A.N-......I-----.......-S.-.-K-PKRI....----------L----------------R------    418
D_ELI       A.--SV....TTA---.......--.-.-K-P--I....----------K-----------R-------L-    425
G_LBV217    A.SGTA....TA----......K-.-.-K-P--N....----D------L---------R-------------    423
H_VI557     -N--.......-VM---......K-.-.-K---RI....----------L--------------------L-    424
O_ANT70     A.QQDLKGGYTAVF--.......--.Q.NPIRKG-....---------------------------Q------    428
O_MVP51     A.QQDLKGGYTAVF--.......--.Q.NPNRKGP....----------K-------R------Q------    428
CPZGAB      -.Q-QG...R-DVFF-......K-.Q.GA-PKRK....----------L----K---R----R--Q------    434
CPZANT      A.N-AQ...GTAVFL-.....--.-GN--GKRP....L----------T----K---R----R--Q----L-    446
A_ROD       -IGP......-P-PFA......AA.Q....---A....F--W-------S--Q-----RQ-------P--I-T    423
B_EHOA      AL-P......STNPFA......AA.Q.P-AGKR-....VT-W----A--T--Q-K---RQ------QQ--I-S    421
SD_MM251    ALAP......VP-PFA......AA.Q.K--P--P....---W-------S--Q-----RQ-------MD-V-A    425
STM_STM     -FQP......DPLPFA......AA.Q.QQ-R-T.....V--W----A--T-KQ-KG--RQ------QQ----A    426
VER_AGM3    M.QS.......QN---......Q-.G.Q--RPRPP...V--Y----F--MQ-Q-PE---MR-L----P--LA-    436
GRI_AGM677  G.Q-.........-V-......V-PQ.KK-P-GP....L-------F--MQ-E-K---QIK-F----I--MA-    424
TAN_AGM17   M.QG.......VN-I-GHSKGG--.R.G--PPR........--K--QI--VQKD-PRGGPIK-L----PR-MA-    434
SAB_SAB1C   AFQQQT...VG--FV-QGA.RP--.P.LG-RGRPLNPN---Y----P--L--F-K---RQ------SPD----    450
SYK_SYK     A.QQ.......SVN-V......Q-.P.SK-RSM.....---Y---QI--MQKD-KK-L-AK-FN---T--LAR    446

                      pol cds ->                                                  VpR
                ->          p7 \/     'p1'      \/ p6                        <-binding->
most-likely DCT.........ERQANFLGKIWPSH.KGRPGNFL.....................QSRP.....EPTAPPA.    459
A_U455      ---.........-------------N.-------P.....................----.....-------.    454
B_HXB2R     ---.........-------------Y.--------.....................----.....------E.    460
C_UG268     ---.........------------PN.--------.....................----.....-------E    455
D_ELI       ---.........--------R-----.--------.....................----.....-------.    461
G_LBV217    ---.........--------------.--------.....................-N--.....--S----.    459
H_VI557     ---.........-----------SPN.--------.....................----.....-------.    460
O_ANT70     --RN........GK-------Y--PG.GT----YV.....................-RPA.....H-S---M.    465
O_MVP51     --KN........G--------Y--PG.GT----YV.....................-KQV.....S-S---M.    465
CPZGAB      ---.........G--V-----G---R.S------V.....................-N-T.....------I.    470
CPZANT      N-PAT......NTGKV-----PT-TWWGC-----V...................QKEEVV.....------I.    488
A_ROD       N-P.........D---G---LGPWG..-K.-R--PVA.................QVPQGL.....T-----V.    461
B_EHOA      K-P.........----G---FGPWG..-K.-R--PVQAPQGIVPSAPPMNPAFGMTPQGA.....I-S----.    476
SD_MM251    K-P.........D---G---LGPWG..-K.-R--PMA.................QVHQGL.....T-----E.    463
STM_STM     K-P.........---VG---FGPWG..-K.-H--PMA.................QIPQGL.....T-----E.    464
VER_AGM3    --RG..........-V----YGRWMG.AK.-R--P...................AATLGV.....------PP    473
GRI_AGM677  --K.........NG------YGHWGG.AK.-R--V.....................-Y-GDTVGL------ME    465
TAN_AGM17   --R.........GG-----RRMPAPX.GSK-R---.....................E-GGA.....V-----IA    472
SAB_SAB1C   --QK..........-V----FGPWGR.GK.-R--P.....................LTSI.....R-----ME    485
SYK_SYK     A-RQPKRNQGPPVA-------G-GVS.RRP-A--P.....................V-S.....--S---LE    491

                                                        \/ minor
most-likely ..........................................ESFGFGEE..TTPSQKQ.EQKDK.........    479
A_U455      ..........................................-I--M--K..M-SPA--.-L--R.........    474
B_HXB2R     ..........................................---RS-V-T.---P---.-PI--.........    481
C_UG268     PTA.................................PPA---R-.-G..---PP--.-P--R.........    480
D_ELI       ..................................................--------..I------.-----.....    481
G_LBV217    ..........................................--L------..IA--P--.-P..-.........    477
H_VI557     ..................................................--------..M--PP--.-R--A.........    480
O_ANT70     ..........................................-EEVK-Q-N.......-.---GGPN........    482
O_MVP51     ..........................................-EAVKEQ-N...Q---G.D-..E.........    482
CPZGAB      ..........................................--Y-YQ--.....EKS-.-K-EG.........    487
CPZANT      ..........................................-IYQE.-HKR-QKGL-G.-E.............    506
A_ROD       ..........................................DPAVDLL-KYMQQGKR-R--RERPYKEVTEDLL    494
B_EHOA      ..........................................DPAEEMLKNYMQLGK--K-NRERPYK.......    502
SD_MM251    ..........................................DPAVDLLKNYMQLGKQ-R-SRE-PYK.......    489
STM_STM     ..........................................DPAADLLRSYMQLGK--R-SRKTPYK.......    490
VER_AGM3    S.....................................PYDPAKKLLQQYADKGKQLR--RK-PPAVN.....    504
GRI_AGM677  T.....................................AYDPAKKLLQQYAEKG-RLR-EREQTRKQ......    495
TAN_AGM17   AHGFPTG.........................PPVAGAYDPAKRLLGQYAKKGDQLRR--E-..........    509
SAB_SAB1C   RDYSRPEENWY................ADRPPTRGPGPDDPATALLKQYAVQGKR-........KQQWQNH.    532
SYK_SYK     DIEDGPWLTWSAQMSQQAQAKAQNSPSKKPPTNREVLSPK--S-K.--T.............-.........    540
```

```
                                VpR
                              <-binding->     / p6 terminus (80%)
most-likely .............E..LYPLASLNSL......KQ                              492
A_U455      .............-..QT--V--K--FGNDPLS-                              493
B_HXB2R     .............-..----T--R--FGNDPSS-                              500
C_UG268     .............-.......P-I--......-S                              488
D_ELI       .............-..----T--K--FGNDPLS-                              500
G_LBV217    .............-..---...-T--......-S                              487
H_VI557     .............-....S..P-T--......-S                              489
O_ANT70     .............-..---F---K--FGT...D-                              498
O_MVP51     .............-..---F---K--FGT...D-                              498
CPZGAB      .............-SS---PT--K--FGSDPSS-                              508
CPZANT      .............-..-P-SY--K--FGK...D-                              522
A_ROD       HLEQGETPYREPPT..EDL-.H----FGK...D-                              522
B_EHOA      ...........EVT..EDL-.H----FGE...D-                              519
SD_MM251    ...........EVT..EDL-.H----FGG...D-                              506
STM_STM     ...........EVT..EDL-.H----FGE...D-                              507
VER_AGM3    .........PDWT-..G-S...----FGE...D-                              521
GRI_AGM677  ..........KEK-V.EDV..--S--FGG...D-                              513
TAN_AGM17   .............-L.EDY.....--......SS                              518
SAB_SAB1C   .....SPQQSPYE-..A-S..--R--FGE...D-                              554
SYK_SYK     .............S..---..--S--FGE...D-                              554
```

```
     **COBBLER sequence from MOTIF**
>GAG   A_ROD, with embedded consensus blocks
mgarnsvlrgkkadelerirlrpggkkkyrlkhivwaankldRFALNPGLLETKEGCQKI
LQQLQPALPTGSENLKSLFNTVCviwcihaeekvkdtegakqivrrhlvaetgtaekmps
tsrptapssekggnypvqhvggnyVHQPLSPRTLNAWVKVVEEKKFGPEVVPMFQALSEG
CTPYDINQMLNCVGDHQGAMQIIKEVINEEAAEWDRTHPipgplpagqlreprgsdiagt
tstveeqiqwmfrpqnPIPVGNIYRRWIILGLQKCVRMYNPVSILDIKQGPKEPFRDYVD
RFYKTLRAEQtdpavknwmtqtllvqnanpdcklvlkglgmnptleemltacqgvgggpgq
karlmaealkevigpapipfaaaqqrkafkcwncgkeghsarqcraprrqgcwkcgkpgh
imtncpdrqagflglgpwgkkprnfpvaqvpqgltptappvdpavdllekymqqgkrqre
qrerpykevtedllhleqgetpyrepptedllhlnslfgkdq
```

```
          p17 ->          /<-       nls        ->/
              /<-    membrane binding     ->/
CONSENSUS-A   mGARaSvLsggkLDawekIrLRPgGkKkYrlKHlvwAsreLerFaLnPslLeTaegcqqimeQlqsalkT     70
CONSENSUS-B   -----------e--r-------------k---i-----------v--g----s---R--lg---ps-q-     70
CONSENSUS-C   ------i-r-----?-----------h-Mi----------------g----s---k--ik---P--Q-     69
CONSENSUS-D   --------------?-------------?---i-------------G----s---k--ig---P-iq-     68
CONSENSUS-F   ----------------------------------------------i--g----s---rk-Ig---pS-Q-     70
CONSENSUS-G   --------------?-----------?--?-----------?-----G----T------??---P?-Q-     63
CONSENSUS-H   -----------?--------------?--------------------?--?-?---L-?I----P----     64
CONSENSUS-O   ---?----T-S------?---?--S--?---------------?-C--?-------?E?LLQ--EP----     62
CONSENSUS-CPZ ---?----?-?---?--??------?-?-M?------??--?---????-?-??---?K???--?P????     42

                                                  /<- nls ->/
CONSENSUS-A   g?eElkSLfNtvatLycvHqrIdvkDtKeAldkiEeiqnKskqk??????tqqaaA..?T.gs?..sskv     126
CONSENSUS-B   -s---r--y-------------e-------------E-----k-......a-----??d-.-n-??--q-     128
CONSENSUS-C   -T---r--?----------??-e-r-----------E---?Q----....----k-.aD?.-k......-     120
CONSENSUS-D   -s------?----------e--e--------e-m--E-----k----...a---t-..D-.rn...--Q-     125
CONSENSUS-F   -S---r--y----v--f----vE----------L--E----q--.....---?--..dK.-.......-     123
CONSENSUS-G   -T--?---?-??-?-?-?----e--------eEV-Ka-kn-Q--.....?---??..e?.-n...--q-     110
CONSENSUS-H   -T---Q----LL-?---------?-------?-??-?--??Q??.....----T?..DK.??...??-?     106
CONSENSUS-O   -S??-?--W-AI?V-W---N-??I?--QQ-IQ-LK-V.M?-RKS...A-AAKE.....--..-...?RQ?     106
CONSENSUS-CPZ ?S????-----??V-W-?-??????--??-????K??????Q??T-S---???G????-????-??????      61

              p17 \/ p24
CONSENSUS-A   ????SqNYPIVQNaqgQm?hQ?lSPrTLnAwVKviEekaFspEVIPmFsaLSEGATpQdLNmMLNiVgGH     190
CONSENSUS-B   ....---------l----V--ai----------v-------------------------T---T----     194
CONSENSUS-C   ....---------L----v--ai-----------------------t-----------t---t----     186
CONSENSUS-D   ....---------L----V--ai---------------------------------------t---T----     191
CONSENSUS-F   ....---------l----V---i-----------------------------------------T---T----     188
CONSENSUS-G   ....-------------------i-----------v-------------------------t---T----     174
CONSENSUS-H   ....-------------?V--AI----------V-------------------------A---?----     170
CONSENSUS-O   ....?------?------V--AI----------AV-----N--I----M------??Y-I-T---AI---     168
CONSENSUS-CPZ ----??---????-?-??----??---------?V---?-?--------------?-?-?-T---A--?-     107

CONSENSUS-A   QAAMQMLKdtINeEAAewDr?HPVhAgPippgQmREPrGSDIAGtTStlqEqigwmTs...NPPiPVGdI     256
CONSENSUS-B   --------e----------l--------a-------------------------n...-------e-     261
CONSENSUS-C   ------------------l-------?a-----------------------a----...---------     252
CONSENSUS-D   --------E----------l--------A----------------------?----...-------e-     257
CONSENSUS-F   -------------------L---q--------i------------------q----...---v---e-     255
CONSENSUS-G   -------------------I--?Q--------I-?----------------R----...-------e-     239
CONSENSUS-H   -----?-------------------------------------------------A---?...--?------     233
CONSENSUS-O   -G-L-V--EV-----?----T--P??--L----I---T----------Q----?-T-R.??-??------     229
CONSENSUS-CPZ -G---V--EV----------L--T----???--L---?----------?---??-??????-???----?     160

                                  /<-     MHR      ->/
CONSENSUS-A   YkrwIiLGLNKIVRMYSPvSILDirQgPKEPFrdYVdrFfKtLRAeqAtQeVKnwMTeTLLvQNANPDCK     326
CONSENSUS-B   ------------------t--------------------Y--------s--------------------     331
CONSENSUS-C   ----------------------k-------------------------D------d--------------     322
CONSENSUS-D   ---------------------------------------Y--------s-d--------------------     327
CONSENSUS-F   ------------------------------------------------------g---D------------     325
CONSENSUS-G   ----------------------------------------------?--?---D------------     307
CONSENSUS-H   ----------------------?----------------?--------?--?---D------------     299
CONSENSUS-O   -RK--V-----M-K----------?--------------Y----------------------?-----     297
CONSENSUS-CPZ -?--?-?----V--?-?-----?-?--------------Y--?-----S-?--?---?---?--------     218

                                                                                Zn-motif
                                        p24 \/  \/     'p2'     \/ p7           /<-
CONSENSUS-A   sILraLg?gAtLeEMMTacQgVggPgHKArvLAEAmSqv...q???n??.iMmQrGnf.rgqkr?iKCFN     384
CONSENSUS-B   T--K---Pa-----------------------------...tn-s.at?--------.-n-rKtv----     394
CONSENSUS-C   T------P--s--------------s------------a...nn.--...-----s--.K-p--iv-----     383
CONSENSUS-D   t--K---P?----------------s------------a...tn.s-ta.---------.K-prki-----     390
CONSENSUS-F   T--K---P-------------------------------a...TN.-?a-.----ks--.K--R-iv-----     386
CONSENSUS-G   T--?---P-----------------?------------A...SG.-A-A.?---K??-.K-P??-----?     360
CONSENSUS-H   ?--?------SI-------------?---?---------?..TN.-?A-.?---K---.K--R-I?----     353
CONSENSUS-O   Q--K?--P?-------V--------T---??-----A?AQQDLKGGYTA.VF----QN.P?R-G------     358
CONSENSUS-CPZ ?--K------?----?---------?----------?????.?Q.-?--.VF?-?-?G??-?---?----     262
```

HIV1 GAG CONSENSUS

```
                                                pol cds ->
                Zn-motif ->/      /<-Zn-motif ->/    p7 \/     'p1'      \/ p6
CONSENSUS-A   CGkEGHlArNCrAPrKkGCwKCgkEGHQmKdCT.?e.rQANFlgkiwpSsKG.RPgNFpQsRp.......    443
CONSENSUS-B   ------i-k-----------------------.--?-----------h--.-----l----???????    453
CONSENSUS-C   ------i----------------?---------..-.-----------?--.-----L----???????    440
CONSENSUS-D   ------i-k-------------------------..-.-----------h--.-----l----.......    449
CONSENSUS-F   ------i-k--------------r----------..-.-----------n--.-----L----.......    445
CONSENSUS-G   ----------------?------?---------..-.-------?----H--.-----L-?-?.......    414
CONSENSUS-H   ------?----------------?----?----..-.----------???--.-----L----.......    406
CONSENSUS-O   ------I-?-------?------Q--------?..NG?-------Y--PGGT.----YV-???.......    411
CONSENSUS-CPZ ------?----K---R----R--Q----?-?-??-????V-----??-???-?-----V-?????.....    306

                    vpr binding                                                vpr binding
                      /<-->/            \/ (minor)        (minor) \/    /<- ->/
CONSENSUS-A   .......EPtAPpAE.......?f?gmgeeit.s?....pkqeqkd..??ke??ppl?slKSlFGNDplS    484
CONSENSUS-B   ??..???------e-.......s-.rf---t-tps????q---pi-...---lY?--a--r-------s-    499
CONSENSUS-C   ???????--------???????S-.rF.--t-.pa....----p--??--?---?--t----------x    480
CONSENSUS-D   .......--------.......S-.-F-----.Ps....q------??----ly.--a-----------    494
CONSENSUS-F   .......--------.......s-.-F?----.PS....--------....egly---a----        482
CONSENSUS-G   .......--?-----.......-?.???---?.?S....----P??....--LY?-------        440
CONSENSUS-H   .......--------.......S-.-F---M-.P-....----??-....?-..?-------        436
CONSENSUS-O   .......?-S---M-.............-?VK.?Q....EN-?--G..--?-LY.-FA-------T-Q$    444
CONSENSUS-CPZ .......------I-........-Y.??Q--?K.?-....?-?????....??L---?--------?-??-    332

                  p6 terminus
                 / (80%)
CONSENSUS-A   Q                                                                       485
CONSENSUS-B   -$                                                                      500
CONSENSUS-D   -                                                                       495
CONSENSUS-CPZ -                                                                       333
```

HMMER Sequences in the Pol Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_HXB2R	HIVHXB2R	K03455	Starcich,B.	Science 227, 484 (1985)
D_ELI	HIVELI	KO3454	Alizon,M.	Cell 46, 63 (1986)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68, 1586 (1994)
O_MVP5180	HIVMVP5180	L20571	Gurtler,L.G.	JVI 68, 1581 (1994)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
CPZANT	SIVCPZANT	U42720	Vanden Haesevelde,M.	Virology 221, 346 (1996)
A_ROD	HIV2ROD	M15390	Clavel,F.	Nature 324, 691 (1986)
B_EHOA	HIV2EHOA	U27200	Rey-Cuille,M.A.	Virology 202, 471 (1994)
SD_MM251	SIVMM251	M19499	Franchini,G.	Nature 328, 539 (1994)
STM_STM	SIVSTM	M83293	Novembre,F.J.	Virology 186, 783 (1992)
VER_AGM3	SIVAGM3	M30931	Baier,M.	Virology 176, 216 (1990)
GRI_AGM677	SIVAGM677	M66437	Fomsgaard,A.	Virology 182, 397 (1991)
TAN_AGMT17	SIVAGMT17	L19254	Hirsch,V.M.	Virology 197, 426 (1993)
SAB_SAB1C	SIVSAB1C	U04005	Jin,M.J.	EMBO J. 13, 2935 (1994)
SYK_SYK	SIVSYK	L06042	Hirsch,V.M.	JVI 67, 1517 (1993)

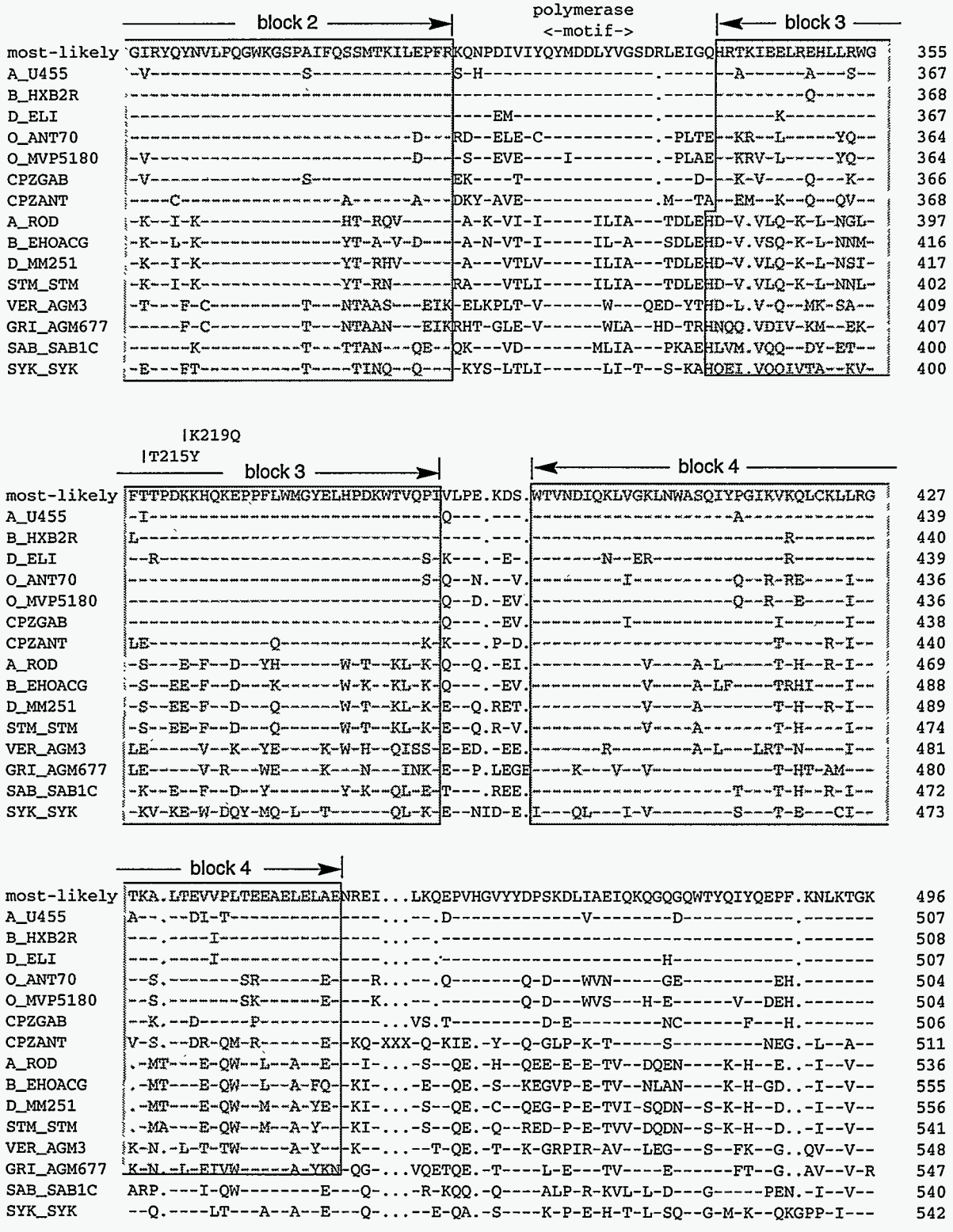
The following alignment and most likely sequence were generated using the HMMER program as described in Part III. For simplicity, only representative types and subtypes are shown. Annotation is based on HIV1s, therefore the user should be cautious about its applicability to other PIV sequences. Five key sites for drug resistance are indicated; for further information regarding drug resistance, see the section by J. Mellors and colleagues in Part III. Cleavage sites are indicated by '\/'. Ten conserved, gapless blocks were identified using the BLOCKMAKER and Motif programs; these are shown by shading. The Motif "Cobbler" sequence follows this alignment.

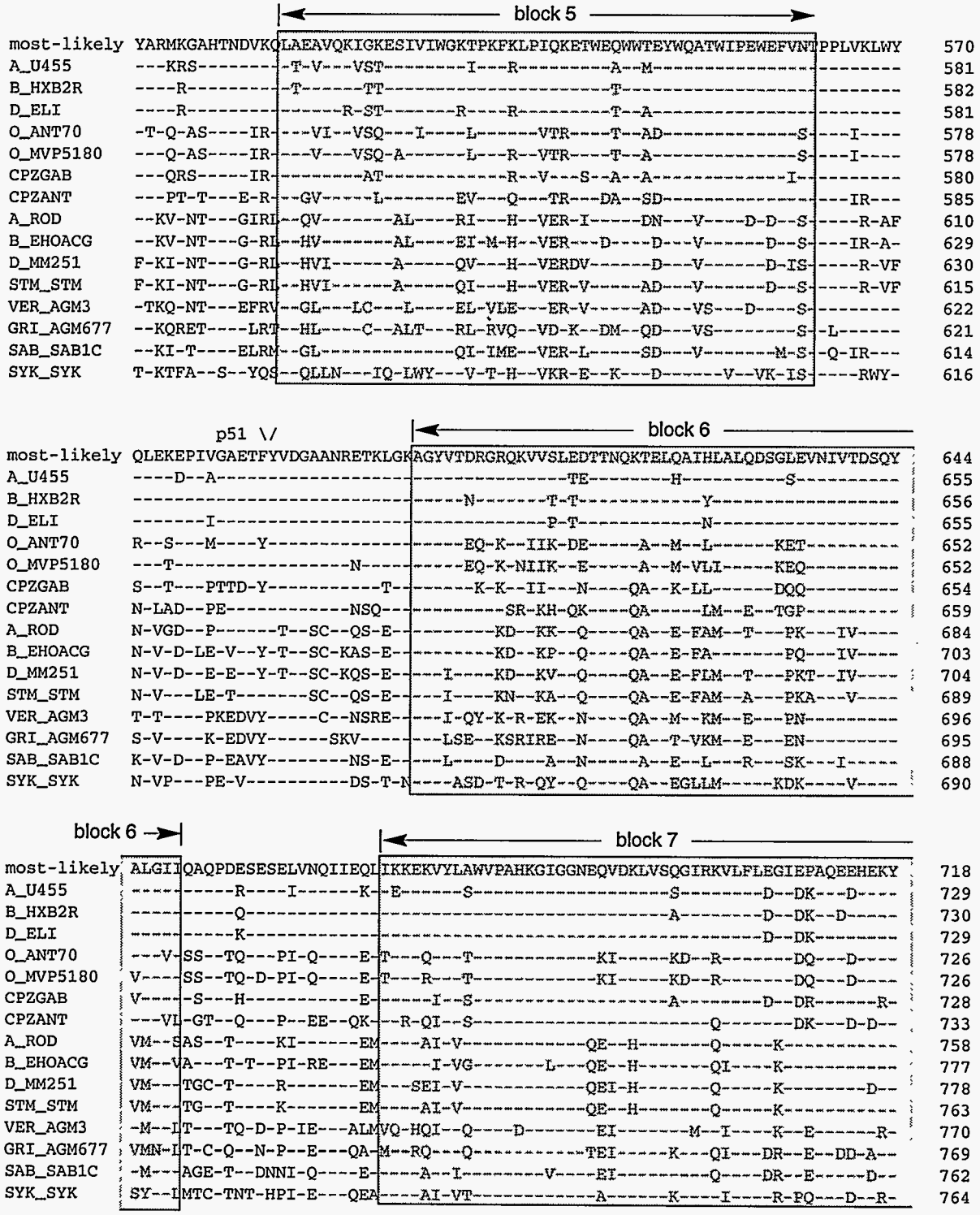
```
most-likely                                                FFRE.DLAFPQGK.AAEQFPR.........     19
A_U455                                                     -----.N---Q--E.-R-F...........     16
B_HXB2R                                                    -----.----L---.-R-F...........     16
D_ELI                                                      -----.N-------.-G-L...........     16
O_ANT70                                                    ---Q.I--S.G-H.E-R-L...........     16
O_MVP5180                                                  -----.V--S.G-H.E-R-L...........     16
CPZGAB                                                     -----.R-----RE.-RQL...........     16
CPZANT                                                     ----T-PHV.V-V.QTRELCA.........     19
A_ROD                                                 TGR---T.GPL...--.E-P-L--..........     19
B_EHOACG       KAGMLEMWTART.................HHVKMPRKTGG---V.RPL...--.E-S----PGTPGDSAI     48
D_MM251        VLELWEGGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGG---P.WSM...--.E-P---H.........     57
STM_STM     RPEKTRLLEMWTTGP.................SDGQMPRKTGG---L.WPM...--.E-P-L-H.........     42
VER_AGM3    T.KKDEMLEMWE.TR.................AFSKRLQRTGK---V.WPVD..-S.ETQK-S-..........     41
GRI_AGM677     VRQNWPYGK.......................RLQEWTGK---V.WPL...-RSETKK-CAIQ.......     36
TAN_AGMT17                                        RRTGQ--K-.NARTLRFQ.--QF-RARR.......     26
SAB_SAB1C                                              ---V.WPLG..QR.ETQE-...........     15
SYK_SYK                                                ----.RMGG.LQE.-PS-L-SEERAICSPS     27

most-likely ..............GPSSSGQARANSPTRRESQ.....................................     38
A_U455      ..................--E-T------S-NLWD.GGKDD.LPCETG.............AER..Q.GT     50
B_HXB2R     ..................--E-T---------L-V.WGRDNNSPSEAG.............ADR..Q.GT     51
D_ELI       ..................-PK-T------S--LRV.WGRDN.PLSKTG.............AER..Q.GT     50
O_ANT70     ..................CAETSTPI---DGGGS..........EGTGES............GTERGPE.RA     47
O_MVP5180   ..................CAETSVPI---NGGGS...EGTRESESEGG.............SGRA.....     47
CPZGAB      ..................CAE-N-T-G--D--LWVPGGR.....EEPGE............ERGR..E.QS     49
CPZANT      ..............EGG---...---S-H-DLS...GGAQ.EDSEGS.............QGG..G.GT     51
A_ROD       ..................----A-.-DT--TPSGS-S...GSTG.EIYAARE..........KTERA.ERET     57
B_EHOACG    ........CAPDEPSIRHDT--.CDSICTPC-S-R...GDAK.ELHATRE..........EAEGE.QRET     94
D_MM251     ..............-S-A--.-D--CSP-GP-C...GSAK.ELHAVGQ............AAERK.QREA     95
STM_STM     ..............--NT--.-D--C-S--P-C...GSAE.KLHAAGQ............EAERE.QEDT     80
VER_AGM3    ..............RY-WG-...--CAPST--I...RPCK.EAPAAI..............CR..QGEA     73
GRI_AGM677  ..............RRH-W--...T---PNGN-L...RSSK.EAPPAVCREGTAPERGERTDKETEG..E.RS     85
TAN_AGMT17  ..............SRSDG-PDR--WI-H-SA-C...RSLRSSQETPGAVCKKR.....RPTKEAKR..E.GT     74
SAB_SAB1C   ..................P-DLHQT--SPNGTGL.........QQAGGK............LVCR..Q.TS     44
SYK_SYK     GGHRRWAMAHMVSTDEPT-T-E..............GTEL.SLEETPHQQG.........SALAEG..E.QW     72

                                                                          <- gag cds end
                                                                \/ protease
most-likely ....................................VGFAFPQITLWQRPLVTAKIEGQLVEALLDTGADD     73
A_U455      ....................................DS-S-------------V--G---I----------     85
B_HXB2R     ....................................-S-N---V---------I--G---K----------     86
D_ELI       ....................................-S-N------------AI--G---K----------     85
O_ANT70     ....................................LSVCL---P--D--I---RVG-H-C-V---------     82
O_MVP5180   ....................................-PICL---P--D--I----VG-H-C-----------     82
CPZGAB      ....................................ISTNL----------IPV-V----C----------     84
CPZANT      ....................................TSLV--E-P-----MMEVL-Q--KCQ---------     86
A_ROD       IQGSDRGLTAPR.............AGGDTIQGATNR-L-A--FS--K--V---Y----P--V---------    115
B_EHOACG    LQG.............................GDR---A--FS--R--V-K-T----S--V---------    134
D_MM251     LQG.............................GDR---A--FS--R--V---H----P--V---------    135
STM_STM     LQG.............................GDR---A--FS--R--V---H----P--V---------    120
VER_AGM3    VEGTKEKTT.....................SSESRLDR-IF-ELP.--R--IK-VY---VPIR---------    123
GRI_AGM677  ....................................G-CFLELP.--R--MKRVI---TP-Q---------    119
TAN_AGMT17  ....................................G-LFLEF                                 81
SAB_SAB1C   DQRTRARRSSNSPVKAVCCSGETAETAVAKPLATTEPLRG-LQL--VS--R--MK-VY----K-T---------    118
SYK_SYK     EGG.............................DQESLSLS-LS--R--MIEVDV--D--QM-V------    112
```

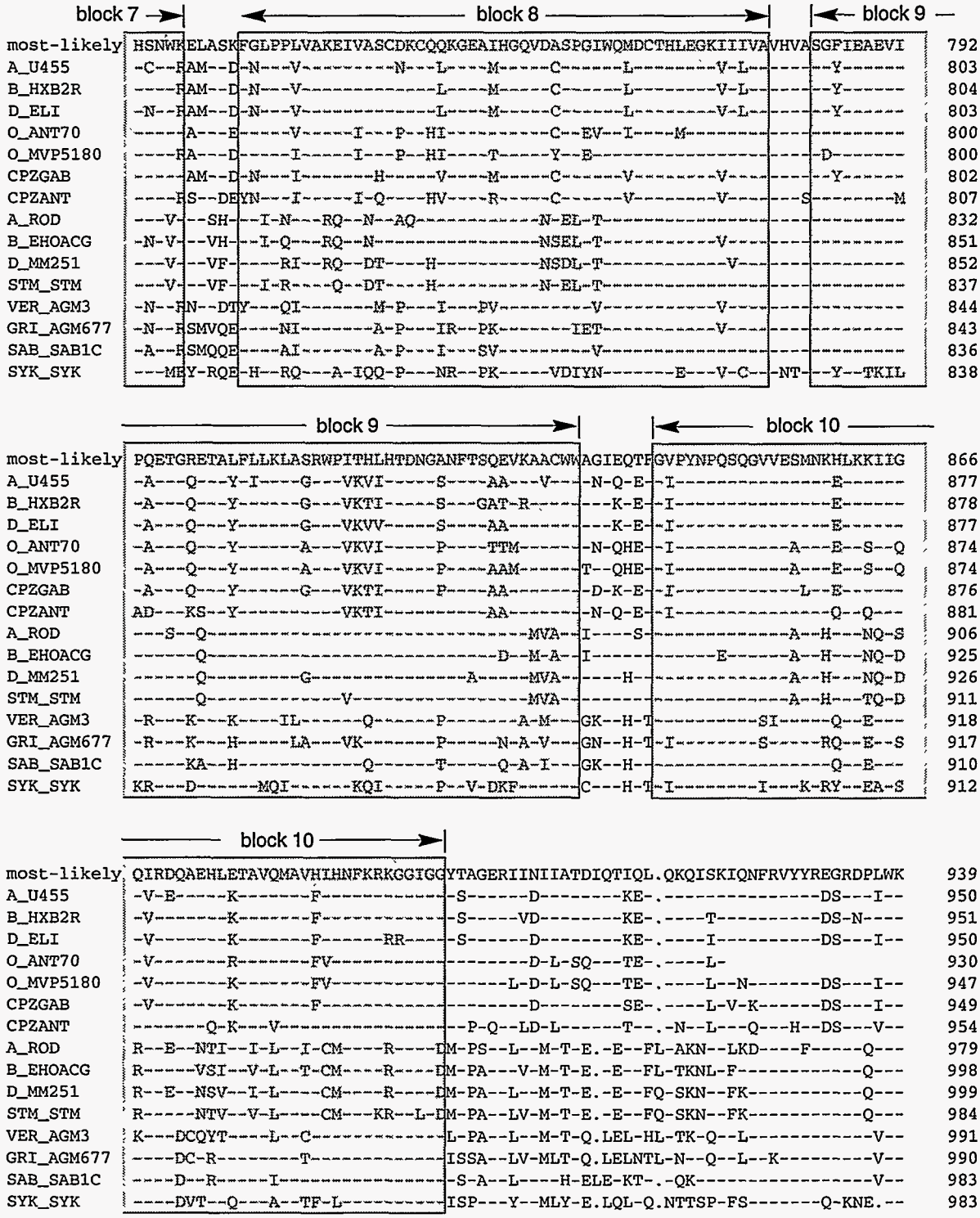
```
most-likely TV..EEGIL..L.GR.WKPKIIGGIGGFIKVKEYDQVEIEIC...G..KRAIGTVLVGPTPVNIIGRNLLTQLG    136
A_U455      --L.-DIN-..P.-K.---------------RQ---IL----...-..-KT----------------M---I-    149
B_HXB2R     --L.--MS-..P.--.----M----------RQ---IL----...-H.-.----------------------I-    150
D_ELI       --L.--MN-..P.-K.----M----------RQ---IP----...-Q.-.----------------------I-    149
O_ANT70     --..LNN-Q..-E-K.----M--------------N-TV--E...-..REVQ----------------I--G--    146
O_MVP5180   --..LNN-Q..-E--.-T--M-------------NN-TV-VQ...-..-EVQ-----------L---I--G--    146
CPZGAB      --..I-R-Q..-Q-L.----M-----------QF-N-H---E...-R.-.VV----------------I-----    148
CPZANT      --..V---H..-Q-N.----T--------S-QQ-NK-P-Q-G...D..RTVLA---L--N--------V-CL--    150
A_ROD       SI..VA--E..-.-NNYS---V--------NT---KN----VL...N..-KVRA-IMT-D--I--F---I--A--    179
B_EHOACG    SI..VA--E..-.-SNYT---V--------NTN--KN----VV...-..--VRA--MT-D--I--F---I-NS--    198
D_MM251     SI..VT--E..-.-PHYT---V--------NT---KN-K--VL...-..--IK--IMT-D--I--F------A--    199
STM_STM     SI..VA--E..-.-LQYT--VV--------NT--FKN-N--VL...-..-KIK--IMT-D--I--F------A--    184
VER_AGM3    -IIK-ADLQ..-S-T.------------GLN----SDR-VRLE...D..-ILR--I-I-S--I------I-APA-    189
GRI_AGM677  -IIQ-KDLHFPP.HKP-RS-VV-----G-H----QG-QVQLE...D..-IIT-SI-I-S--I------I-A-A-    187
SAB_SAB1C   S-..IQ--E..-.-DN---R--------C-N--A-HNQ-VK-E...D..-TCKA-I---E---------V-A---    182
SYK_SYK     -IIR--D-Q..-.HQP-S---V--L--N-T-RQ-RNIRFTVVKPS-KR-QVE--L---------L---I--K--    183
```

```
      protease \/ p66,p51                                          |M41L          <-- block 1 --
most-likely CTLNFPISP..IETVP.VKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPVFAIKKK    207
A_U455      ---------..-----.-----E-------------------T---N-------------------------------    220
B_HXB2R     ---------..-----.-------------------------------------------------------------    221
D_ELI       ---------..-----.-------------------------T----D--------R----------I-------    220
O_ANT70     ---------..-AP--.-----------------SK---E--TA--Q---Q-----R----------I-------    217
O_MVP5180   ---------..-AP--.-----------------SR---E--TA--Q---Q-----R----------I-------    217
CPZGAB      ---V----S..-----.-----------------SA------T---Q--------------------I-------    219
CPZANT      --------K..V----.----E-----R------SK---E--K---DKL-A-N---R---D------I-------    221
A_ROD       MS--L-VAK..V-PIK.IM----K----LR-----K---E--K---EK-----QLEEAP-T------T-------    250
B_EHOACG    M-----VAR..--P-K.-Q---EK----IR----SK---L--K---EK-----QLEEAP-T----S-T-------    269
D_MM251     MS--L--AK..V-P-K.-T----KV---L-----SK---V--R---EK---D-QLEEAP-T------T-------    270
STM_STM     MS----VAK..V-P-K.-T----K----I-----SK---E--K---EK---D-QLEEAP-T------T-------    255
VER_AGM3    AK-VMGQLSE.QIPITP----E-AR--FL-----SK------Q---DQL----------G--A------C-----    262
GRI_AGM677  MK-VMGVLSSQ--ETK.-Q--E-K----L-----SR---E--T---KQ--E---L-R--G-----------------    260
SAB_SAB1C   V---LTQRE..--PIK.-H----Q---RIR----SK---E--KA--EDL--Q-HLER----------R--        253
SYK_SYK     VK-VMVQTA..L-P-K.-S---DKEL-RL-----SV--LE--KA-VED-L-A-QLE-AS-T--------V-R--    254
```

```
   |K70R
  |D67N
                  block 1 ------------------>                                  <-- block 2 --
most-likely DSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTAFTIPSVNNETP    281
A_U455      ----------------------------T----------|----------S----------I-----            294
B_HXB2R     ------------------------------------------|----------|----------I-----        295
D_ELI       ------------------------------------------|----------|----------S-I-----      294
O_ANT70     -G----------------E----------G---Q-Q-----|---------C---P----------------      291
O_MVP5180   ------------------------------G---QRQ----|---------C---P----------------      291
CPZGAB      ------------------------------------------|--------C---K----------I-----      293
CPZANT      -TS-------------------I----------Q--------|---------I---Q----------------      295
A_ROD       -KN---M-I-------V----T-I----------A--RRI--|---------I--H----P-----L-----AE-    324
B_EHOACG    -KN---M-I-------V--E-T----------AS--RI----|----------P----Q-----L-A---AE-      343
D_MM251     -KN---M-I-------RV----T----------A-R-RI---|--I------I----E--Q-----L-----AE-    344
STM_STM     -KN---M-I-------V----T-I----------A-RRRI--|---------I----G--Q-----L-----AE-    329
VER_AGM3    -KSQ--M---------A----F---------S-FE-MTEI--|--I----Y-I---PE----------QG-       336
GRI_AGM677  -K-Q--M---------A----F----------Q---QI---|-I----Y-I--CKE----------TG-         334
SAB_SAB1C   -K-Q--I-M---Q---S----Q----------QQREQI---|--I------C---P--Q----------RE-      327
SYK_SYK     -KK---M-I---K--EA----F---------G--.-RQKL-II-LK---Y-----KE--P-----V--I--AS-    327
```


block 2
polymerase
<-motif->
block 3
block 4
K219Q
T215Y
most-likely
A_U455
B_HXB2R
D_ELI
O_ANT70
O_MVP5180
CPZGAB
CPZANT
A_ROD
B_EHOACG
D_MM251
STM_STM
VER_AGM3
GRI_AGM677
SAB_SAB1C
SYK_SYK


block 5
block 6
p51 \/
block 6
block 7


block 7
block 8
block 9
block 9
block 10
block 10
most-likely HSNWKELASKFGLPPLVAKEIVASCDKCQQKGEAIHGQVDASPGIWQMDCTHLEGKIIIVAVHVASGFIEAEVI 792
A_U455 -C--RAM--D-N---V----------N---L----M-----C------L---------V-L-----Y------- 803
B_HXB2R -----RAM--D-N---V-------------L----M-----C------L---------V-L-----Y------- 804
D_ELI -N--RAM--D-N---V-------------L----M-----C------L---------V-L-----Y------- 803
O_ANT70 -----A---E-----V-----I---P--HI----------C--EV--I----M---------------------- 800
O_MVP5180 ----RA---D-----I-----I---P--HI----T-----Y--E-------------------D------- 800
CPZGAB -----AM--D-N---I--------H------V----M-----C------V---------V-----Y------- 802
CPZANT ----RS--DEYN---I-----I-Q----HV----R-----C------V---------V-------S--------M 807
A_ROD ---V---SH---I-N---RQ--N--AQ----------N-EL-T---------------------------- 832
B_EHOACG -N-V---VH---I-Q---RQ--N----------------NSEL-T----------V---------- 851
D_MM251 ---V---VF-----RI--RQ--DT----H----------NSDL-T----------V---------- 852
STM_STM ---V---VF---I-R----Q--DT----H----------N-EL-T---------------------- 837
VER_AGM3 -N--RN--DTY---QI--------M-P---I---PV---------V----------V---------- 844
GRI_AGM677 -N--RSMVQE----NI--------A-P---IR--PK-------IET----------V---------- 843
SAB_SAB1C -A--RSMQQE----AI--------A-P---I---SV---------V---------- 836
SYK_SYK ---MEY-RQE-H--RQ---A-IQQ-P---NR--PK-----VDIYN-------E---V-C--NT---Y--TKIL 838
most-likely PQETGRETALFLLKLASRWPITHLHTDNGANFTSQEVKAACWWAGIEQTFGVPYNPQSQGVVESMNKHLKKIIG 866
A_U455 -A---Q---Y-I----G---VKVI-----S----AA---V---N-Q-E--I----------E------ 877
B_HXB2R -A---Q---Y------G---VKTI-----S---GAT-R-----K-E--I----------E------ 878
D_ELI -A---Q---Y------G---VKVV-----S----AA---------K-E--I----------E------ 877
O_ANT70 -A---Q---Y------A---VKVI-----P----TTM------N-QHE--I----------A---E--S--Q 874
O_MVP5180 -A---Q---Y------A---VKVI-----P----AAM------T--QHE--I----------A---E--S--Q 874
CPZGAB -A---Q---Y------G---VKTI-----P----AA-------D-K-E--I----------L--E------ 876
CPZANT AD---KS--Y----------VKTI----------AA-------N-Q-E--I----------Q--Q--- 881
A_ROD ---S--Q----------------------------MVA--I----S----------A--H---NQ-S 906
B_EHOACG ------Q----------------------D--M-A--I----------E------A--H---NQ-D 925
D_MM251 ------Q---------G----------A------MVA--------H----------A--H---NQ-D 926
STM_STM ------Q----------V----------MVA--------------------A--H---TQ-D 911
VER_AGM3 -R---K---K----IL------Q------P-------A-M---GK--H-T----------SI-----Q--E--- 918
GRI_AGM677 -R---K---H-----LA---VK-------P------N-A-V---GN--H-T-I--------S-----RQ--E--S 917
SAB_SAB1C -----KA--H----------Q------T-----Q-A-I---GK--H-----------Q--E--- 910
SYK_SYK KR---D------MQI------KQI-----P--V-DKF------C---H-T-I---------I---K-RY--EA-S 912
most-likely QIRDQAEHLETAVQMAVHIHNFKRKGGIGGYTAGERIINIIATDIQTIQL.QKQISKIQNFRVYYREGRDPLWK 939
A_U455 -V-E-----K-------F----------S------D--------KE-.---------DS---I-- 950
B_HXB2R -V-------K-------F----------S-----VD--------KE-.----T----------DS-N---- 951
D_ELI -V-------K-------F-------RR----S------D--------KE-.----I----------DS---I-- 950
O_ANT70 -V-------R-------FV----------------D-L-SQ---TE-.----L- 930
O_MVP5180 -V-------K-------FV----------------L-D-L-SQ---TE-.----L--N-------DS---I-- 947
CPZGAB -V-------K-------F----------------D--------SE-.----L-V-K------DS---I-- 949
CPZANT -------Q-K---V----------------P-Q--LD-L------T--.-N--L---Q---H--DS---V-- 954
A_ROD R--E--NTI--I-L--I-CM----R----DM-PS--L--M-T-E.-E--FL-AKN--LKD----F-----Q--- 979
B_EHOACG R-----VSI--V-L--T-CM----R----DM-PA---V-M-T-E.-E--FL-TKNL-F----------Q--- 998
D_MM251 R--E--NSV--I-L----CM----R----DM-PA--L--M-T-E.-E--FQ-SKN--FK----------Q--- 999
STM_STM R-----NTV--V-L----CM---KR--L-DM-PA--LV-M-T-E.-E--FQ-SKN--FK----------Q--- 984
VER_AGM3 K---DCQYT----L--C----------L-PA--L--M-T-Q.LEL-HL-TK-Q--L----------V-- 991
GRI_AGM677 ----DC-R---------T----------ISSA--LV-MLT-Q.LELNTL-N--Q--L--K---------V-- 990
SAB_SAB1C ----D--R-----I----------S-A--L----H-ELE-KT-.-QK----------V-- 983
SYK_SYK ----DVT--Q---A--TF-L---------ISP---Y--MLY-E.LQL-Q.NTTSP-FS-------Q-KNE.-- 983

```
                                        vif cds ->
most-likely GPAKLLWKGEGAVVIQ.DGSDIKVVPRRKAKIIRDYGKQMAGDDCV.............D                  985
A_U455      ----------------.-N--------------------------M...AGRQD..E.-                 1002
B_HXB2R     ----------------.-N---------------------------...ASRQD..E.-                 1003
D_ELI       ----------------.-K----------V----------------...ASRQD..E.-                 1002
O_MVP5180   ---Q-------------.-KG---------------------T-SM...ANRQT..E.SESMEQPGEIP       1009
CPZGAB      ---T-------------.-QGEL-------------------------...ASRQN..E.-KHGTAW         1007
CPZANT      ---Q-----------                                                             969
A_ROD       --GE---------LVK.V-T---II------------GRQEM-SGSHLEGAREDG.EMA                 1036
B_EHOACG    --GD---------I-K.V-TE---I---------N--GGKEL-CSADVEDTMQAR.EVAQSN              1058
D_MM251     --GE---------ILK.V-T-------------K---GGKEV-SSSHMEDTGEAR.EVA                 1056
STM_STM     --GE---------IVK.V-T-------------K---GGKEV-SGSHLEDTREAG.EVA                 1041
VER_AGM3    --GQ-I---------K.G-VEL-EY--------K--EPRKRMG-ESNLEGAGGA..D.N                1046
GRI_AGM677  ---R-I---------K.E-E-------------K---ERKTM-SEGSMEGVREANKQMEGDSDLQDQE       1057
SAB_SAB1C   -----I----------.EQGEL-TI--------K----ALDSQAPL.EGNGRTAG.EV-                1039
SYK_SYK     ---R----------V-TEEG--FA---------T-H-ER-DSGSH-.....ENDP.KT-                1036
```

COBBLER sequence from MOTIF

```
>POL   CPZGAB, with embedded consensus blocks
ffrerlafpqrearqlcaeqnrtngptdrelwvpggreepgeergreqsistnlpqitlw
qrplipvkvegqlcealldtgaddtvieriqlqglwkpkmiggiggfikvkqfdnvhiei
egrkvvgtvlvgptpvniigrniltqlgctlvfpissietvpvklkpgmdgpkvkqwpls
aekikalteicqemekegkiskigpeNPYNTPVFAIKKKDKTKWRMLVDFRELNKRTQDF
WEVQLGIPHPAGLKKKKSVTVldvgdayfscpldkDFRKYTAFTIPSVNNETPGVRYQYN
VLPQGWKGSPAIFQSTMTKILEPFReknpditiyqymddlyvgsdleidqHRRMVEQLRE
HLLKWGFETPDKKHQKEPPFQWMGYELHPDKWTIQKIqlpekevWTVNDIQKLVGKLNWA
SQIYPGIKTKHLCKLIRGTKSLTEVVPWTEEAEAELEEnreivstpvhgvyydpdkelia
eiqkqgncqwtyqifqephknlktgkyarqrsahtndirqLAEVVQKIGKESIVIWGKIP
KFHLPVERETWEQWWTDYWQATWIPEWEFVSTpplvklwysletepipttdtyyvdgaan
retktgkAGYVTDRGKQKVKKLEQTTNQQAELEAILMALQDSGPKVNIVTDSQYAMGIIq
sqpdhseselvnqiieelIKKEKIYLAWVPAHKGIGGNEEVDKLVSQGIRQVLFLEGIDE
AQEEHEKYHSNWRamasdFGLPPIVAKEIVASCPKCQIKGEAMHGQVDCSPGVWQMDCTH
LEGKVIIVAvhvaSGFIEAEVIPQETGQETAYFLLKLASRWPVKHIHTDNGPNFTSQEVK
AACWWadikqefGIPYNPQSQGVVESMNKHLKEIIGQIRDQAEHLETAVLMAVHIHNFKR
KGGIGGytageriidiiatdiqtselqkqilkvqkfrvyyrdsrdpiwkgpatllwkgeg
avviqdqgelkvvprrkakiirdygkqmagddcvasrqnedkhgtaw
```

HIV1 POL CONSENSUS

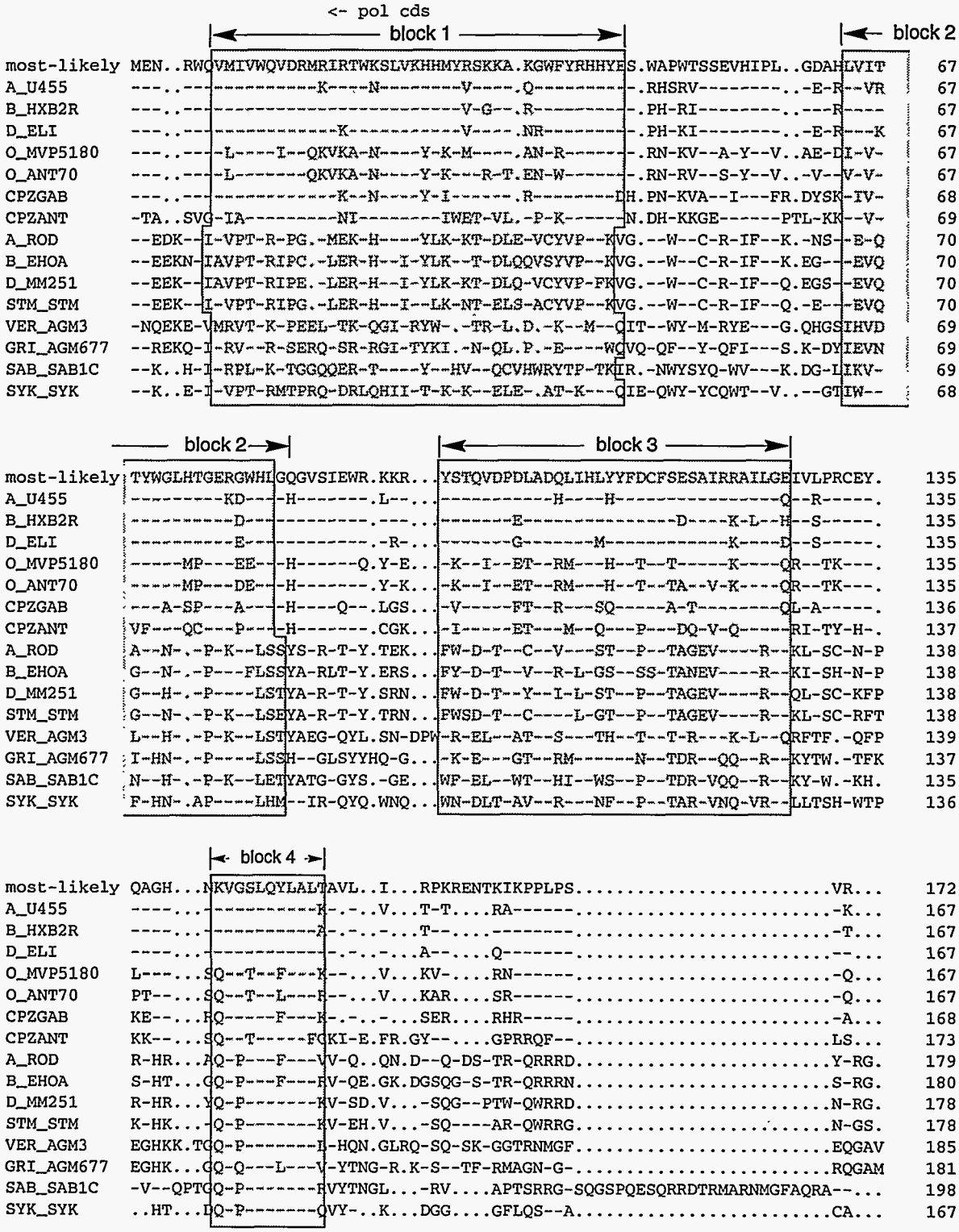
```
CONSENSUS-A   FFRE.NLAFQQGEAR?F.............SSE..QT??NS?TSR?LWDGG?D??.L?...???G?E?..Q     35
CONSENSUS-B   -----.d---p--k--e-????????????---..--Ra--p-r-E-qVw-r-nnS-S???-EA-adr..-     49
CONSENSUS-D   -----.d---P--K-GEl.............---..--RA--P---E-RVW-r-.NP-S....eT-A-R..-     48
CONSENSUS-O   ---?.?--SGGH---QL.............CA-..TS-PI-P-?.....--GSE.....GT-ES?---G??     35
CONSENSUS-U   ----.----P--K--E-.............P--..--RA--P---E-RVW-G-K.T-S....ET-A-R..-     48
CONSENSUS-CPZ ----???????????--L.............CA-????---?--?---????-?--??-.....--?-?-???     13

                    protease
                      \/           <- gag cds end
CONSENSUS-A   G?..??SF?FPQITLWQRPLVTV?I?GQLIEALLDTGADDTVLEDINLPGKWKPK?IGGIGGFIKVRQYD     96
CONSENSUS-B   -...tV--s-------------ik-g---K--------------eM----r----M--------------    116
CONSENSUS-D   -...TV--n-------------IK-G---K--------------Em---------M--------------    115
CONSENSUS-O   R...A-??CL---P--D--I--A-VG-H-C-?-----------NN-Q-E-?-?--M----------KE-?     94
CONSENSUS-U   -...IV--S----------V---RVG---K--------------E----------M--------------    115
CONSENSUS-CPZ ?-??---?-?-?-?-----???--?---?C?-----------?-?-?-Q-?-------------?-?-??     55

                                          protease \/ p66, p51
CONSENSUS-A   QILIEICGKK?IGTVLVGPTPVNIIGRNMLTQIGCTLNFPISPIETVPVKLKP?MDGPKVKQWPLTEEKI    164
CONSENSUS-B   --------H-A-----------------L------------------------G----------------    186
CONSENSUS-D   --------?-A-----------------L------------------------G----------------    184
CONSENSUS-O   NVTV-??-?EVQ------------?---I--GL-----------AP-------G----------S?---    159
CONSENSUS-U   ----------A---I-------------------------------------G----R----------    185
CONSENSUS-CPZ ?V?-?-??R?V??---?--?--------?-??L----?----??--------?G----?------S?---    106

                    | M41L                D67N |   | K70R
CONSENSUS-A   KALT?IC?EMEKEGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPH?AGLK    231
CONSENSUS-B   ---vE--T-------------------------------------------------------P----    256
CONSENSUS-D   ----E--T---------R----------I-----------------------------------P----    254
CONSENSUS-O   E---A--Q---Q-----R----------I-------?----------------?----------PG---    227
CONSENSUS-U   ----E--KD-------L--------------------------N-------------------P----    255
CONSENSUS-CPZ ?--?E---??-?-?---?---?------I-------??--------------------?------P----    164

CONSENSUS-A   KKKSVTVLDVGDAYFSVPLD??FRKYTAFTIPS?NNETPG?RYQYNVLPQGWKGSP?IFQ?SMTKILEPF    295
CONSENSUS-B   --------------------kd-----------i------i---------------A---s---------    326
CONSENSUS-D   --------------------eD-----------I------I---------------A---S---------    324
CONSENSUS-O   Q?Q-------------C---PD-----------V----------------------A---S------D--    295
CONSENSUS-U   --------------------ED-----------I------I---------------A---S---------    325
CONSENSUS-CPZ ?---------------?----D--------------------?-----------------------?--    225

                  polymerase motif
                       /<-->/                    T215Y |    | K219Q
CONSENSUS-A   R???P?IVIYQYMDDLYVGSDLEIGQHRAKIEELR?HLL?WGF?TPDKKHQKEPPFLWMGYELHPDKWTV    358
CONSENSUS-B   -KQN-d----------------------t------q---r---t-------------------------    396
CONSENSUS-D   -KQN-E----------------------T------E---R---T-------------------------    394
CONSENSUS-O   ---N-E?E-?--?----------PL?E--KR?-L--E--YQ---T-------------------------    358
CONSENSUS-U   -TKN-E----------------------T------E---K---T-------------------------    395
CONSENSUS-CPZ --K---??-------------?--??--???-?--Q--?---?-------------?------------    278

CONSENSUS-A   QP??LPEKDSWTVNDIQKLVGKLNWASQIYAGIK?KQLC?LLRGAKALTDIV?LTEEAELELAENREI..    421
CONSENSUS-B   --Iv-----------------------------V----k----t----Evip---------------..    464
CONSENSUS-D   -sIk----E----------------------p---Vr---K----T----EViP---------------..    462
CONSENSUS-O   -?IQ--?-?V---------?----------Q--RV?E--K-I--T-S--EV-P-S?------E----?..    419
CONSENSUS-U   --IQ--D-E----------------------P---V----K------------P--A------------..    463
CONSENSUS-CPZ -?I----???---------?----------P----------I--?-?---?--?-?------?--??-??    329

CONSENSUS-A   .LK?PVHGVYYDP?KDLVAE?QKQGQDQWTYQIYQEPFKNLKTGKYA?KRSAHTNDVKQLTEVVQKV??E    484
CONSENSUS-B   .--e---------s---i--i-----g--------------------rm-G----------A---iat-    533
CONSENSUS-D   .--E---------S---I--i----hG--------------------Rm-G---------a-a---IsT-    531
CONSENSUS-O   .----------Q-D---WV?I--?-??-----?--?EH--------?RQKAS----IR--A--?---SQ-    479
CONSENSUS-U   .--E---------S---I--I-----G---------QY---------RIK-----------A---IAQ-    532
CONSENSUS-CPZ ???-???-?--?--??-?-?I----??------?-???-?--?----R????---??R--A??---I---    367
```

```
                                                                  p51 \/
CONSENSUS-A   SIVIWGK?PKFRLPIQ?ETWE?WWMEYWQATWIPEWEFVNTPPLVKLWYQLEKDPI?GAETFYVDGAANR    550
CONSENSUS-B   -------t---k----K-------t----------------------------e--v-------------    602
CONSENSUS-D   -------T--------K----T--?----------------------------E--I-------------    600
CONSENSUS-O   ?-?----L---?--VTR----T--A?-------------S----I----?--?E-------?--------    541
CONSENSUS-U   -------T--------K----A--T----------------------------TE--V-------------    602
CONSENSUS-CPZ ------?----?--??--?-?A--??------------?-----??---?-???--P???-?--------    416

CONSENSUS-A   ETK?GKAGYVTDRGRQKVVSLTETTNQKTELHAIHLALQDSGSEVNIVTDSQYALGIIQAQPDRSESE?V    618
CONSENSUS-B   ---l-----------------d--------q----------l--------------------k----l-    672
CONSENSUS-D   ---L---------------Pf-D--------Q--N-------L--------------------K----L-    670
CONSENSUS-O   ?--L--------EQ-K-?IIK-?------A--M-?L?-----KE?---------?--?-SS--TQ-?-PI-    602
CONSENSUS-U   ---K-------------------------Q--------------------------------K----I-    672
CONSENSUS-CPZ ???---------?-??????-??----QA--?-?L?--?--???---------?--??-??--?---?L-    459

CONSENSUS-A   NQIIEKLI?K?KVYLSWVPAHKGIGGNEQVDKLVS?GIRKVLFLDGIDKAQE?HE?YH?NW?AMASDFNL    681
CONSENSUS-B   s----q--K-E----a------------------a----------------e--K--s--r--------    742
CONSENSUS-D   s----Q--K-E----A------------------Q----------------E--K--N--R--------    740
CONSENSUS-O   Q----E-TK-E?---T------------KI-----KD--R----E---Q---D--K--S----L--?-G-    669
CONSENSUS-U   -----Q--Q-D-------------------------S----------------E--K--S--R--------    742
CONSENSUS-CPZ ??--??--K?E?I-------------------------?----?---?-----?---S---??-???--    510

CONSENSUS-A   PP?VAKEIVASC?KCQ?KGEAMHGQVDC?PGIWQLDCTHLEGK?ILVAVHVASGYIEAEVIPAETGQETA    746
CONSENSUS-B   --v---------d---L----------S--------------v--------------------------    812
CONSENSUS-D   --V---------D---L----------S--------------V--------------------------    810
CONSENSUS-O   --------I---P--HI----?-----?S-E?--?----?---I-I-------?F---------------    732
CONSENSUS-U   --I---------D---L----------S--------------I-I------------------------    812
CONSENSUS-CPZ --I-----?-?-D--?V----?------S-----V--------V-I-----?--?-----???---??--    569

CONSENSUS-A   YFILKLAGRWPVKVIHTDNGSNFTSAAVKA?CWWAN??QEFGIPYNPQSQGVVESMNKELKKIIGQVR?Q    812
CONSENSUS-B   --l----------t----------tT---A----GiK-----------------------------D-    882
CONSENSUS-D   --l----------V----------------A----GIK-----------------------------D-    880
CONSENSUS-O   --L----A------------P----??M--A---??IQH---------------A------S--Q---D-    798
CONSENSUS-U   ---------------V----------------A-----IK-----------------------------E-    882
CONSENSUS-CPZ --L----?-----T------?---------A----?I-----------------?--?--?----?-D-    631

CONSENSUS-A   AEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIIDIIA?DIQTKELQKQI?KIQNFRVYYRDSRDPIWKGP    880
CONSENSUS-B   ---------------------------------v----t----------T---------------l----    952
CONSENSUS-D   ------------------------------------T----------i--------------------    950
CONSENSUS-O   ----?--------V------------T----?---L-SQ---T------L-?N-----------------    865
CONSENSUS-U   ----------------------------------M--T----------T----------N--------    952
CONSENSUS-CPZ --?-----?---?-------------T?-?--?--?-T----??--?--L-?-?---?-------?----    687

                                  vif cds ->
CONSENSUS-A   AKLLWKGEGAVVIQDNSDIKVVPRRKAKIIRDYGKQMAGDDC?AGRQDED                       929
CONSENSUS-B   ----------------------------------------V-s-----                       1002
CONSENSUS-D   -------------------------V--------------V-S-----                       1000
CONSENSUS-O   -Q-------------KG---------------------T-SM-N--T-SESMEQPGEIP            925
CONSENSUS-U   ------------------------------------------V--G----KHGTAW              1008
CONSENSUS-CPZ -?-------------QGEL----------------------V-S--N--KHGTAW               742
```

HMMER Sequences in the Vif Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_HXB2R	HIVHXB2R	K03455	Starcich,B.	Science 227, 538 (1985)
D_ELI	HIVELI	K03454	Alizon,M.	Cell 46, 63 (1986)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68,1586 (1994)
O_MVP5180	HIVMVP5180	L20571	Gurtler,L.G.	JVI 68, 1581 (1994)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
CPZANT	SIVCPZANT	U42720	Vanden Haesevelde,M.	Virology 221, 346 (1996)
A_ROD	HIV2ROD	M15390	Clavel,F.	Nature 324, 691 (1986)
B_EHOA	HIV2EHOA	U27200	Rey-Cuille,M.A.	Virology 202, 471 (1994)
SD_MM251	SIVMM251	M19499	Franchini,G.	Nature 328, 539 (1994)
STM_STM	SIVSTM	M83293	Novembre,F.J.	Virology 186, 783 (1992)
VER_AGM3	SIVAGM3	M30931	Baier,M.	Virology 176, 216 (1990)
GRI_AGM677	SIVAGM677	M66437	Fomsgaard,A.	Virology 182, 397 (1991)
SAB_SAB1C	SIVSAB1C	U04005	Jin,M.J.	EMBO J. 13, 2935 (1994)
SYK_SYK	SIVSYK	L06042	Hirsch,V.M.	JVI 67, 1517 (1993)

The following alignment and most likely sequence were generated using the HMMER program as described in Part III. For simplicity, only representative PIV types and subtypes are shown. The BLOCKMAKER and Gibbs programs were employed to identify four conserved, gapless blocks (shown by shading); the Motif program did not give satisfactory results in this case. The Gibbs' "Cobbler" sequence follows this alignment.



```
most-likely .KL.TEDSRWNKQRG.............................SEPQKTKGHR.G..SH.TMN.....GH     203
A_U455      .--.---.----.....................................----R---.-..-R.-L-.....R-     192
B_HXB2R     .--.---.----.....................................--------.-..--.---.....--     192
D_ELI       .--.---.----.....................................--Q-R---.-..--.---.....--     192
O_MVP5180   .R-.---.----.....................................-W-IRDQL.-..--.S--.....--     192
O_ANT70     .--.---.----.....................................HLRIRDQL.K..-P.S--.....--     192
CPZGAB      .--.---.----.....................................H-R--V-Q.E..NL.-R-.....--     193
CPZANT      .I-.---.----.....................................-RRMR---.E..NQ               193
A_ROD       LR-AKQ---SH---S..............................--SPTPRTYFP-V.AEVLEI.....LA     215
B_EHOA      IRMARDNI-TSQ-SS..............................-QSLAQGTYFP-L.AEVLGI.....LA     216
D_MM251     LRMAKQN--GD----..............................-K-PTKGADFP-L.AKVLGI.....LA     214
STM_STM     IRVA-QNG-GH-P--..............................-K-STEGTDFP-L.AKVLGI.....LA     214
VER_AGM3    GRMAKRHA-RYQSGSQDAFWA.....................RAPVP-MELLSG-G-KESH--.ARK.....-L     232
GRI_AGM677  GRMA-RHAQGS-R-SQKAL.....................WNEHAN-SMELLC-.-..GKE-             219
SAB_SAB1C   .RMAPRHVTGPQF--PVPLP.....................KESPFPSLVEYCG-.T..--XS..WANQA-R     243
SYK_SYK     .RN.-MVLHSK-C-VDPKRDQCHCKGRTGSDRSIQAFYSSRNIWSL-SILKRRG-D                    221
```

COBBLER sequence from GIBBS

>VIF CPZGAB, with embedded consensus blocks

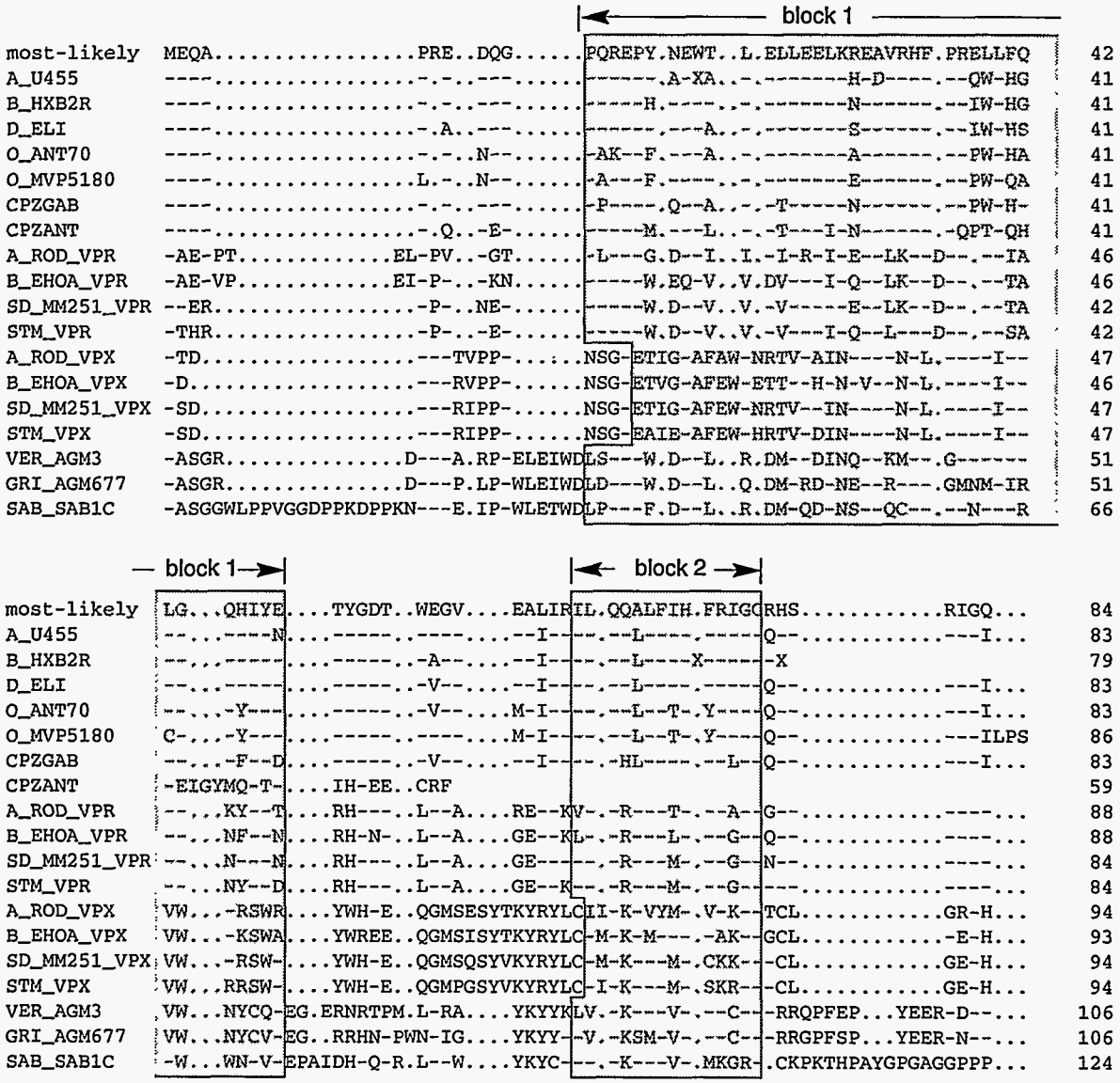
menrwqVMVVWQVDPRRIERWHSLVKYHMYKSKKLWKWRYHHHYEhpnpkvaseihipfr
dyskLVVTTYWGLTPGEGWWHLghgvsiqwrlgsYWTDVDPDTADRLIHSYYFPCFTESA
VRRAIRGElvaprceykeghrQVGSLQYLALKaliserrhrpplpsvakltedrwnkhqr
tkvhqenltrngh

```
                                  <- pol cds
CONSENSUS-A    MENRW.Q.VMIVWQVDRMrIRTWNSLVKHHMYVSKKAkGWFYRHHfEsRHpkvsSEVHIPLGd..ARLVV     66
CONSENSUS-B    -----?-.---------------k--------i-g----------Y--t--ri----------..-----i     66
CONSENSUS-D    -----.-.---------------K-------------?R------Yd-p---I---------E..-----     65
CONSENSUS-O    -----.-.-L----?--QKVKA------Y-K-?-?-??N-?----Y---N-?---?-Y--V??..-??--     54
CONSENSUS-CPZ  -????.?.-??---------??-?----?-I???-????-?----Y????--???-?---??????K-?-     34

CONSENSUS-A    RTYWGLHTGErDWHLGhGVSIEWrqKRYSTQvDPDLADqLIHLhYFdCFSdSAIRkAILGeiVRPRCEYQ    136
CONSENSUS-B    t---------------q-------k------------------y------e----n----h--s------    136
CONSENSUS-D    k----------?----Q-------KR--------G-------MY------E?--------h--S?-----    132
CONSENSUS-O    T-----MP---?E----------?Y-?-K--I--ET--RM------T--T?--?------QR-LTK---?    118
CONSENSUS-CPZ  T??-?-??---?---------?--?G?-?-----?T--??--??--?--???-?-?----??????-?-K     76

                                                 vpr cds ->
CONSENSUS-A    AGHNKVGSLQYLAL.kAL...VaPtkaKPPLPSvkKLtEDRWnePQKTRGHRGsR?mNgH$              191
CONSENSUS-B    --------------.a--...it-k-i-------?--------K----K-----ht-----              191
CONSENSUS-D    ---?----------.t--...i--K-I-------R--------K----k--?--HT----              186
CONSENSUS-O    ?--SQ--T--?---.?-V...-K????-------Q?-------K???I-DQL?-?S-----             161
CONSENSUS-CPZ  ?--?Q--?--?--??-?-?????????R???--???-------K??R???-?EN?TR---             107
```

HMMER Sequences in the VPR-VPX Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_HXB2R	HIVHXB2R	K03455	Starcich,B.	Science 227, 538 (1985)
D_ELI	HIVELI	K03454	Alizon,M.	Cell 46, 63 (1986)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68, 1586 (1994)
O_MVP5180	HIVMVP5180	L20571	Gurtler,L.G.	JVI 68, 1581 (1994)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
CPZANT	SIVCPZANT	U42720	Vanden Haesevelde,M.	Virology 221, 346 (1996)
VER_AGM3	SIVAGM3	M30931	Baier,M.	Virology 176, 216 (1990)
GRI_AGM677	SIVAGM677	M66437	Fomsgaard,A.	Virology 182, 397 (1991)
SAB_SAB1C	SIVSAB1C	U04005	Jin,M.J.	EMBO J. 13, 2935 (1994)

The following alignment and most likely sequence were generated using the HMMER program as described in Part III. For simplicity, only representative PIV type and subtype sequences are shown. Because annotation cannot simultaneously apply to VpR and to VpX, see the accompanying HIV-1 consensus alignments for this. Two conserved, gapless blocks were found using the BLOCKMAKER and Motif (or Gibbs) programs; these are shown by shading. The Motif "Cobbler" sequence follows the alignment.



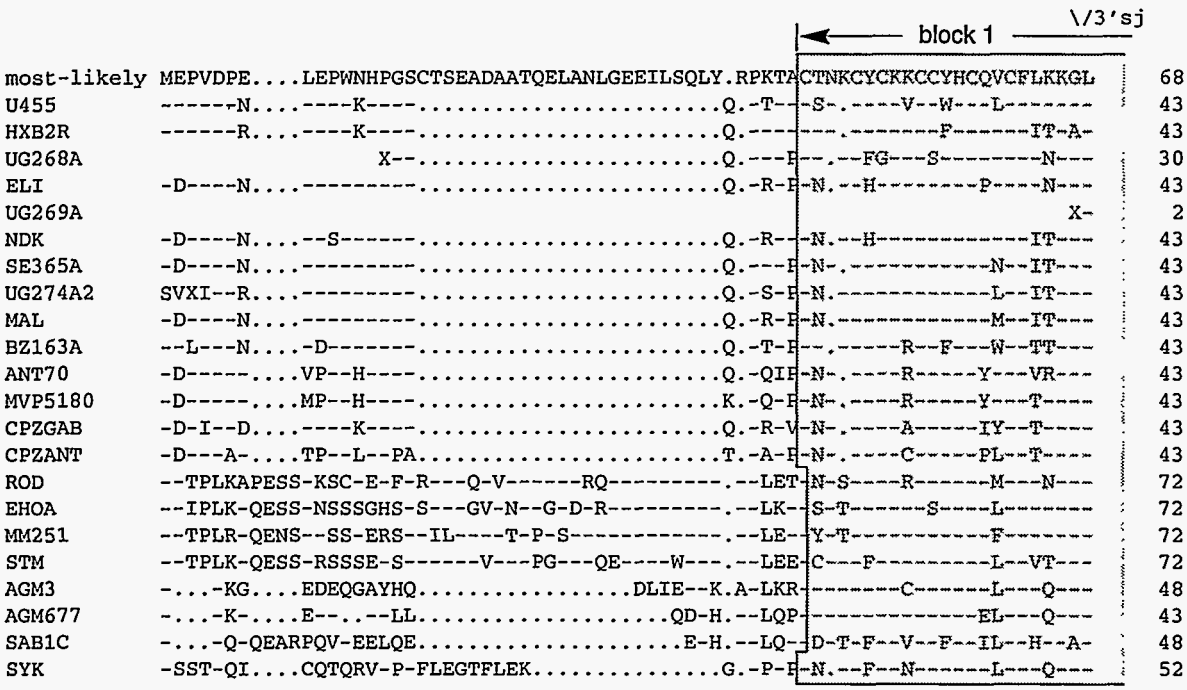
```
most-likely   .G.GG.RRGPPPRNPPSRS                    100
A_U455        .I.P-.--...G--GA---                     96
D_ELI         .I.RQ.--...A--GS---                     96
O_ANT70       .NPR-.-G...R--GS---                     97
O_MVP5180     NT.R-.-G...R--GS---                    100
CPZGAB        .L.PQ.--...RS-GSN--                     96
A_ROD_VPR     .T.R-.GN.-LSAI-TP-NMQ                  105
B_EHOA_VPR    .P.--.GN.-LSAI---X                     102
SD_MM251_VPR .P.--.GN.-LSTI---                        97
STM_VPR       .P.--.GN.-LATI--T-GVL                  101
A_ROD_VPX     .-P--W-P----PP--GLV                    112
B_EHOA_VPX    .-P--W-S----PP--GLA.X                  112
SD_MM251_VPX .-A--W-P----PP--GLA                     112
STM_VPX       .-A--W-P----PP--GLA                    112
VER_AGM3      .-.--.-A...G-V--GLD                    119
GRI_AGM677    .-.--.AP....PP--GL                     117
SAB_SAB1C     .-L--.AS-GAASAA-GL                     140

      **COBBLER sequence from MOTIF**
>VPR_VPX   D_ELI, with embedded consensus blocks
meqapadqgPQREPWDEWLLEMLEELNQEAVRHFPRELLFQLWQYIYEtygdtwvgveai
irILQKALFMHFRxGCqhsrigiirqrrarngssrs
```

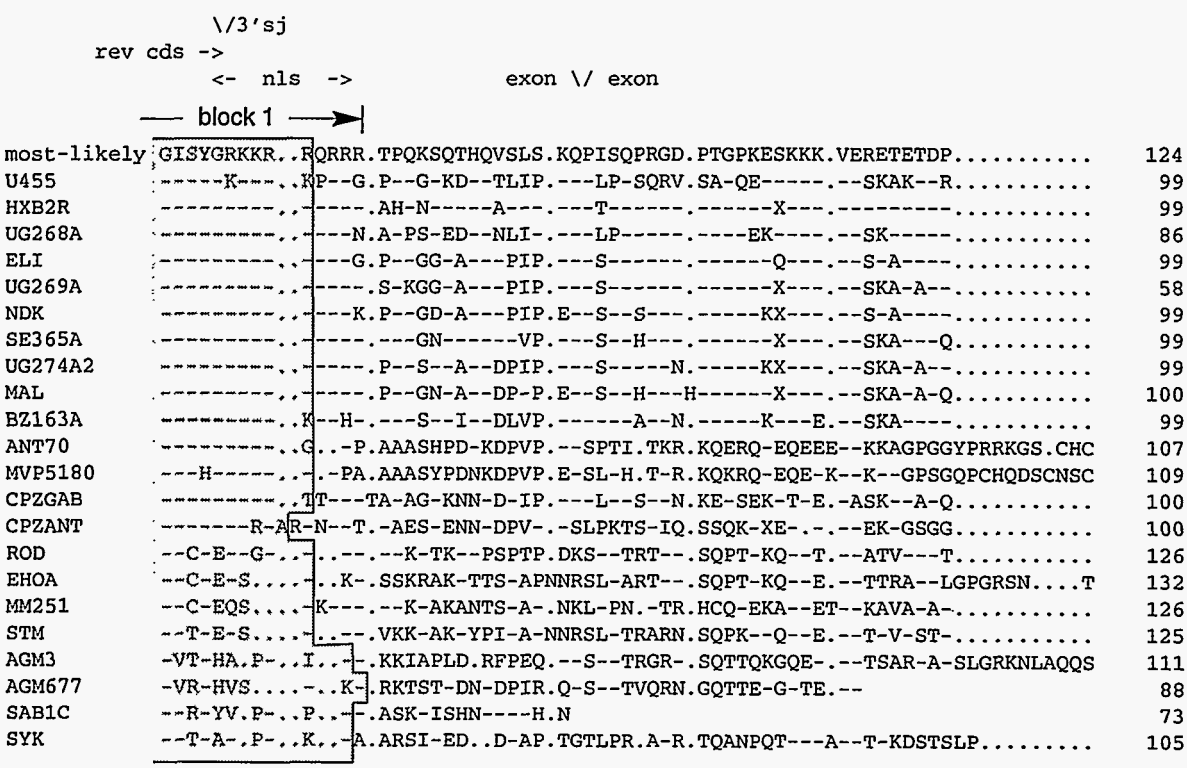
```
                                  <- vif cds                                  LR domain
                 /<-          oligomerization          ->/                    /<-
CONSENSUS-A      ME?..AP.EDQGPQREP??E??LELLEELKHE?VRHFPR?WLHGLGQHIY?TYGDTWEGV?AIIRILQQL      58
CONSENSUS-B      --q??--?---------yN-Wt--------?-A------i---?------E------a--E---------      65
CONSENSUS-D      --Q..--.---------YN-Wt--------S-A------I---S---?--E------?--E-?-------      64
CONSENSUS-O      --Q..--.-n---a---fN-Wt--------?-A------p---a---y--E---------m---------      66
CONSENSUS-U      --Q..--.A--------HN-WT--------Q-A------I---S------E---------E----S----      67
CONSENSUS-CPZ    --Q..--.?-?--?------W---T---?-N-A-----?P?-????-???-?-???????-??????-??      33

                     LR domain ->/  tat cds ->
CONSENSUS-A      LF?H.FRIGCQHSRIGII...?GRRG.RNGA?RS$                                       84
CONSENSUS-B      --i-?-----r------t...-q--a?----S---                                       93
CONSENSUS-D      --I-.------------t...RQ--A.---SS--                                        93
CONSENSUS-O      --t-.y-----------????-rg--r---SS--                                        94
CONSENSUS-U      --I-.------------T...RQ--A.---SS--                                        96
CONSENSUS-CPZ    ??I-.????-??-----L...PQ--R.S--SN--                                        54
```

HMMER Sequences in the Tat Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_HXB2R	HIVHXB2R	K03455	Starcich,B.	Science 227, 484 (1985)
C_UG268A2	HIVUG268A2	L22948	Louwagie,J.J.	JVI69, 263 (1995)
D_ELI	HIVELI	K03454	Alizon,M.	Cell 46, 63 (1986)
F_BZ163A	HIV1BZ163A	L22085	Louwagie,J.J.	ARHR 10, 561 (1994)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68, 1586 (1994)
O_MVP5180	HIVMVP5180	L20571	Gurtler,L.G.	JVI 68, 1581 (1994)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
CPZANT	SIVCPZANT	U42720	Vanden Haesevelde,M.	Virology 221, 346 (1996)
A_ROD	HIV2ROD	M15390	Clavel,F.	Nature 324, 691 (1986)
B_EHOA	HIV2EHOA	U27200	Rey-Cuille,M.A.	Virology 202, 471 (1994)
SD_MM251	SIVMM251	M19499	Franchini,G.	Nature 328, 539 (1994)
STM_STM	SIVSTM	M83293	Novembre,F.J.	Virology 186, 783 (1992)
VER_AGM3	SIVAGM3	M30931	Baier,M.	Virology 176, 216 (1990)
GRI_AGM677	SIVAGM677	M66437	Fomsgaard,A.	Virology 182, 397 (1991)
SAB_SAB1C	SIVSAB1C	U04005	Jin,M.J.	EMBO J. 13, 2935 (1994)
SYK_SYK	SIVSYK	L06042	Hirsch,V.M.	JVI 67, 1517 (1993)

The following alignment and most likely sequence were generated by the HMMER program as described in Part III. For simplicity, only representative types and subtypes are shown. The annotation is based on HIV1s, therefore the user should be cautious about its applicability to other PIV sequences. The BLOCKMAKER program was employed to identify one conserved, gapless block (using either the Motif or Gibbs approach); the block has been indicated with shading. A putative nuclear localization signal (nls) is shown. The Motif "Cobbler" sequence follows this alignment. For further information regarding the Tat protein, see the section by K.-T. Jeang in Part III of this compendium.





```
most-likely ....G.DR                                                    127
U455        ....F..A                                                    101
HXB2R       ....F.-X                                                    102
UG268A      ....X.-X                                                     89
ELI         ....XDCX                                                    103
UG269A      ....F.-X                                                     61
NDK         ....FDWX                                                    103
SE365A      ....F.-W                                                    102
UG274A2     ....F.-W                                                    102
MAL         ....FDWX                                                    104
BZ163A      ....XDXX                                                    103
ANT70       CTRTS.EQ                                                    114
MVP5180     TRIS-.QX                                                    116
CPZGAB      ....X.-X                                                    103
CPZANT      ....-.PC                                                    103
ROD         ....-PG-                                                    130
EHOA        STSRFANX                                                    140
MM251       ....-LG-                                                    130
STM         ....-LGK                                                    129
AGM3        GRAT-.AS                                                    118
SYK         ..SAE.NL                                                    110
```

```
      **COBBLER sequence from MOTIF**
>TAT   UG274A2, with embedded consensus blocks
svxidprlepwnhpgsqpstpCNKCYCKKCCYHCQLCFLKKGLGISYGRKKRRqrrrppq
ssqahqdpipkqpssqprgnptgpkkxkkkveskaeadpfdw
```

```
                                  intramolecular    3'sj   3'sj
                                 disulfide bonding   \/     \/
                                 |                 | rev cds ->/<- nls ->/
CONSENSUS-A   M?PVDPnLEPWnHPGSqPtTaCskCYCK?CCwHCqlCFLnKGLGISYGrKKR..r?RRgtPQs?kDhQnp     64
CONSENSUS-B   -e----r----k------k---tn----k--f---v--tt------------..-Q--ra--dSqt--vs     68
CONSENSUS-C   ------?----------K---t-----k-sY--lV--qt------------..-q--sa-?-SE-----     65
CONSENSUS-D   -d----------------?-p-N--h--K--Y---v--it------------..-Q--rp--ggQa--?-     66
CONSENSUS-F   -EL-----D----------P-T-----R--F---W--TT------------..KQ-HR----SQI--DL     68
CONSENSUS-O   -D----E?P--H----?-Q?P-NN----R--Y--YV--??------?-----...-???AAA--P-?KD-     55
CONSENSUS-U   -D----K----------K---T-----K--Y--PV----------------..-P--RS--NSE-----     68
CONSENSUS-CPZ -D-?-????--?--???-?-?-NN-------Y--??--TK----------?-???---T????S?NN-D?     45

            exon \/ exon
CONSENSUS-A   ipKQplPqtqg??ptgpkESkKkVeSKteTDrf?$                                      95
CONSENSUS-B   Ls---?s-pr-D.------------rE----P?d?                                      99
CONSENSUS-C   -s-------r-d.----E------------p-D-                                       98
CONSENSUS-D   -----SS-pR-d-------?-------A---p-Dw$                                      99
CONSENSUS-F   V----IS-AR-N.-----?---E----A??-P?--$                                      96
CONSENSUS-O   V-?-S???-?RK.Q?RQE-QE??--K??GP?G?P????SC??CTR?S?Q$                       83
CONSENSUS-U   ----S--H--RV.S---E----E----A-----D-                                      101
CONSENSUS-CPZ ??-??-?????-.?????K??-?-??--?????-?                                      52
```

HMMER Sequences in the Rev Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_HXB2R	HIVHXB2R	K03455	Starcich,B.	Science 227, 484 (1985)
D_HIVELI	HIVELI	K03454	Alizon,M.	Cell 46, 63 (1986)
F_BZ163A	HIV1BZ163A	L22085	Louwagie,J.J.	ARHR 10, 561 (1994)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68, 1586 (1994)
O_MVP5180	HIVMVP5180	L20571	Gurtler,L.G.	JVI 68, 1581 (1994)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
CPZANT	SIVCPZANT	U42720	Vanden Haesevelde,M.	Virology 221, 346 (1996)
A_ROD	HIV2ROD	M15390	Clavel,F.	Nature 324, 691 (1986)
B_EHOA	HIV2EHOA	U27200	Rey-Cuille,M.A.	Virology 202, 471 (1994)
SD_MM251	SIVMM251	M19499	Franchini,G.	Nature 328, 539 (1994)
STM_STM	SIVSTM	M83293	Novembre,F.J.	Virology 186, 783 (1992)
VER_AGM3	SIVAGM3	M30931	Baier,M.	Virology 176, 216 (1990)
GRI_AGM677	SIVAGM677	M66437	Fomsgaard,A.	Virology 182, 397 (1991)
SAB_SAB1C	SIVSAB1C	U04005	Jin,M.J.	EMBO J. 13, 2935 (1994)
SYK_SYK	SIVSYK	L06042	Hirsch,V.M.	JVI 67, 1517 (1993)

The following alignment and most-likely sequence were generated using the HMMER program as described in Part III. For simplicity, only representative types and subtypes are shown. The annotation is based on HIV1s, therefore the user should be cautious about its applicability to other PIV sequences. The high affinity binding site to RRE, which also contains a nuclear localization signal (nls) and the leucine effector domain are shown. The BLOCKMAKER program (using the Motif or Gibbs approach) produced one conserved, gapless block, which is shown by shading. The Motif "Cobbler" sequence follows this alignment.

```
           \/3'sj                exon\/ exon        <-   nls       ->

                                 |<-------------- block 1 ----------------
most-likely MAGRSG..DSDEILKTVRLIKLLYQS......NPYPSPEGTRQARRNRRRRW.RERQRQIRSISARILSTFL     64
A_U455      --R----..NP--D---A--I-------.......--C-N-R-S----K-------.-A-----D-L-E----DC-     64
B_HXB2R     -------.------IR-----------.......--P-N--------------.-------H---E---G-Y-     64
D_ELI       -------.-----D---A-----F----.......--P----------------.-A------E-AE---G-Y-     64
F_BZ163A    -------.N-GT--RAA-Y--I----.......----K--------------.-A------E--R----SC-     64
O_ANT70     ----....SE-DQ--QAIQI--I----.......--Q---R-S-N--K------.-R--A-VDTLA--V-A-VV     62
O_MVP5180   -----...EE-QQ--QAIQI--I----.......--C-T-A-S-N--K------.-R--A-VD-LAT---A-VV     63
CPZGAB      -----EP.QD-AR--QA-KI--I----.......----------K---------.-A--K--SE--G-V-A-Y-     65
CPZANT      ----EELEGT--Q---A-KI--I---.......----K-A-S-A----K----.KK--D-VEGLA--V-R-L      64
A_ROD       -NE-....ADE-G-QRKL---R--H-T.......----QGP--ASQ--------.KQ-W---L.AL-DSIY--.     60
B_EHOA      -NA-.......ERD-Q-GL--LH--H-TSEYGTTD---QGP--ASQ--------.KQ-GL--L.AL-DRIHPL.     64
D_MM251     -..S-H..ERE--R-RL---H--H-T.......S--TGP--ANQ--Q-----.-R-WQ-LL.AL-DRIYS-.     60
STM_STM     -...-...-QE--R-RL---QF-H-TT.....D---QGP--ANQ--------.-Q-WN-LL.AL-NRIYS-.     59
VER_AGM3    -...P...LGS--.RRLL---AF-NKN.......----PV---ARQ--RA-----.-QA-E-L-.AL-ERIWHS.     57
GRI_AGM677  -...-...LGK--.KQALKI--T--G-.......----QFS--ARQ--RA-Q--.-KQ-Q--DK-AG-V-N--.     58
SAB_SAB1C   -...-...LGQ--.--RRF-I--F--TT.......----PGQ--ARQ--RA-Q--AKQ--QV-H.LAE---....     56
SYK_SYK     -...P...-QGS--.QLA-F-RMIAHLQ......E---G---P--T--R---Q-.-Q-RT-RL.YLQQRI....     54

                      leucine-rich
                       <-domain->
           block 1 -->|
most-likely GRSAEPVP.LQLPPLERLTLDCSE...DCGTSGT.Q.......GVGSP.......QILVE.............    105
A_U455      --P-----.-----I---R-----...S------.-QPQGTET---G-........--S--.............    112
B_HXB2R     --------.-------------N-...--------.-.......-----.....-----.............    105
D_ELI       --P-----.---------N-N---...--R----.-.........---H-.........--S--.............    105
F_BZ163A    --PE----.-----P----INN--...N-EQGAE.E.......----------TSG-.............    105
O_ANT70     HGPQNNNI.VD-----Q-SIRDP-...GDQL-EA.W.......T-DPRAEDNXCL-N-CS.............    110
O_MVP5180   HG-QDNNL.VD-----Q-NIRDP-...ADRLP--.G.......T-DPG.......TKDN              103
CPZGAB      --PPK-GD.-E--E-DK-S-Q-V-TTQ-V---N-S-.......P...Q.......TATG-TVPA.........    111
A_ROD       ...PD-.-A..DS--DQT.IQH.L...QGL-IQE.L.......P.DP-.......TH-P-.............     93
B_EHOA      ...PDS.-T..EG--DLA.IQR.L...QNLIIKD.L.......P.NP-.......TSTPTAQASTCIPPIWDQ    110
D_MM251     ...PD-.-T..DT--DLA.IQQ.L...QNLAIES.I.......P.DP-.......TNTP-AL............     95
STM_STM     ...PD-.-A..ST--DLA.VQQ.L...QGLSIQD.L.......P.DP-.......PN-PK.............     92
VER_AGM3    .-VE-Q.L.V-A..IDQ-V--Q.....QHLAIQQ.L.......P.DP-.......SSS               89
GRI_AGM677  ...EDQQLVA--QE-QLENK-LVL...QHL.............P.DP-.......H-H               89
SAB_SAB1C   ...ET--S.QID.H-AQEFDQLVL...-NL..QQ.P.......PSLP-.......GHPT-.............     91
SYK_SYK     ...F-A.I.FGS..RTAA.-ED-L...QQLQISD                                        77

most-likely ..........................SPTVLESG...........TK..E                        116
A_U455      ..........................-SA--G--............--..N                        123
B_HXB2R     ..............................--------.............--..-                    116
D_ELI       ..............................--------.............-EEQC                    118
F_BZ163A    ..........................-HA--G--............--..-                        116
O_ANT70     ........................CN-I-ATR...........IA..-                          121
CPZGAB      ........................GGNYSI-GK-...........A-..N                        124
A_ROD       .........................-Q.R-..............AE..T                          100
B_EHOA      LVPRSNPSSSQGCGRDSCERGEDLVG--..Q---RRDHCNTQEDQ-R..G                        156
D_MM251     ..........................CD--KGSRS...........PQ..D                        107
STM_STM     ............................D..-QDT...........AE..N                        100
SAB_SAB1C   ........................NQ-ANS-S                                          99
```

COBBLER sequence from MOTIF

>REV ELI, with embedded consensus blocks

magrsgdsdedLLRAIRLIKLLYQSNPYPQPEGTRQQRRNRRRRWRQRQRQIxALAERIL

STylgrpaepvplqlpplerlnlncsedcrtsgtqgvghpqisvesptvlesgteeqc

HIV1 REV CONSENSUS

```
                                                  high-affinity
                                                  binding site
                                                      nls
              \/ 3' sj              exon \/ exon    /<-          ->/
CONSENSUS-A   MAgRSG?sDE.eLL.KAiRIIKiLYQSNPyPkPkG.SRQARKNRRRRWRARQRQIDSlSeRILStCLGRP     66
CONSENSUS-B   ------d---.---.-tV-l--f------p-s-e-.T----R-------e-----r-i--w----y---s     67
CONSENSUS-F   ------N-?T.---.R-?-Y-------------E-.T----R-----------?-R??-?----S-----     61
CONSENSUS-O   -----E-.-..Q?-?Q--Q----------?-?-?-.--N----------R--A-V-?-A?-?-A-VVHG?     56
CONSENSUS-U   ------DA--.---.RVV-------------P-E-.T--T---------------RAI---F-------S     67
CONSENSUS-CPZ ----?E-??????-??-VK-------?----?-?-.?-?--R-?----??--?-??????-V-?-??---     41

                    Leu-rich
                 effector domain
                   /<-    ->/
CONSENSUS-A   AEPVPLQLPPlERLhLDCsEdcgTSgTQq?qg?etGVGrpQvsVEssavLGSGTkn                  120
CONSENSUS-B   --------------t---?---------......?---s--il---p---e----E$                 115
CONSENSUS-F   E---------?---?IN?--?-E.Q-A?E.....---S--T-G--H--------E$                 105
CONSENSUS-O   Q?NN?VD-----Q-?IRDP-?D?L????TVDPRAEDN$CL-NLCSCNT????????N$                 95
CONSENSUS-U   ----------I---C-----G--------P--T-----S-PI-G---TI------E$                 123
CONSENSUS-CPZ PK-GD-E--E-DK-S-Q-V-TTQDV--SNTSQPQ-AT-ETVPAGGNYSI--K-A--                  97
```

HMMER Sequences in the Vpu Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_SF2	HIVSF2	K02007	Sanchez-Pescador,R.	Science 227, 484 (1985)
D_ELI	HIVELI	K03454	Alizon,M.	Cell 46, 63 (1986)
F_BZ163A	HIV1BZ163A	L22085	Louwagie,J.J.	ARHR 10, 561 (1994)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68, 1586 (1994)
O_MVP5180	HIVMVP5180	L20571	Gurtler,L.G.	JVI 68, 1581 (1994)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
CPZANT	SIVCPZANT	U42720	Vanden Haesevelde,M.	Virology 221, 346 (1996)

HIV1 VPU

The following alignment and most likely sequence were generated using the HMMER program as described in Part III. For simplicity, only representative HIV1s are shown. Two sites of phosphorylation are indicated. One conserved, gapless block was found using the BLOCKMAKER and Motif (or Gibbs) program, which is shown by shading. The "Cobbler" sequence from the Motif analysis follows this alignment. For further information about the VpU protein, see the section by K. Strebel in Part III of this compendium.

```
                                                                  env cds ->
                                                                  phos |     |phos
                                                                       <-- block 1 -->
most-likely MQPLQILAIVALVVALIIAIVVWTIVFI...EYRKILR...QRKIDRL.IDRIRERA..EDS.GNES.EGDQE   62
A_U455      -T--E-W--TG-I----L--------G-...-$K-NC$S..KK-$TG#.LN------..---.----.D--T-   59
B_SF2       --S-------S---VA----------L-...-------...--#----.------K-..---.----.-----   61
D_ELI       ----G-I--A-----I-L----------...---R-KK...--R--C-.L---T---..---.----.---R-   62
F_BZ163A    -SD-LAISVT--I------------Y-...----L--...----N--.YEG-----..---.----.---A-   62
O_ANT70     -HHRDL---IIISAL-F-NVIL-GFILRKYL-QKEQD-K..E-E-LER.LR----IR..D--.DY--.N-EE-   66
O_MVP5180   -HQENL--LI--SALCL-NVLI-LFNLRIYLVQ--QD-R..EQE-LER.LR--K-IR..D--.DY--N-EE-Q  67
CPZGAB      -TL-VG-VL.I--GLIAWN-CI-GYIIKW..G--RYK-HRLETE-E--.NLIL----..---.----.N-EE-   65
CPZANT      -TNIFEY-F......-AFS--L-I-CIPILYKLY--YK...-QQ--NKRNQ--I-VLSRRL-IDSAI.-E-E-   63

                                                                                         81
    block 1 -->|                                                                        78
most-likely .E.LSALVEM.GHHAPWDVD...DL                                                    80
A_U455      .-.--L----.-NYDLGVDN...N-                                                    81
B_SF2       .-.-------.--L------...--                                                    81
D_ELI       .K.--K----.-------I-...--                                                    85
F_BZ163A    .-.-A--G--.-PFI-G-IV...N-                                                    85
O_ANT70     Q-.VMD--LSH-FDN-MFEP.....                                                    88
O_MVP5180   .-.VME-IHSH-FAN-MFEL.....                                                    83
CPZGAB      .-R-EQ-IHN.YN-NNHFANPMF--
CPZANT      .A.DTYYLGS.-FAN-VYREG..-E
```

```
     **COBBLER sequence from MOTIF**
>VPU   A_U455, with embedded consensus blocks
mtpleiwaitglivalilaivvwtivgiekkncskkktglnRIREVAEDSGNESNGDEEE
lsllvemgnydlgvdnnl
```

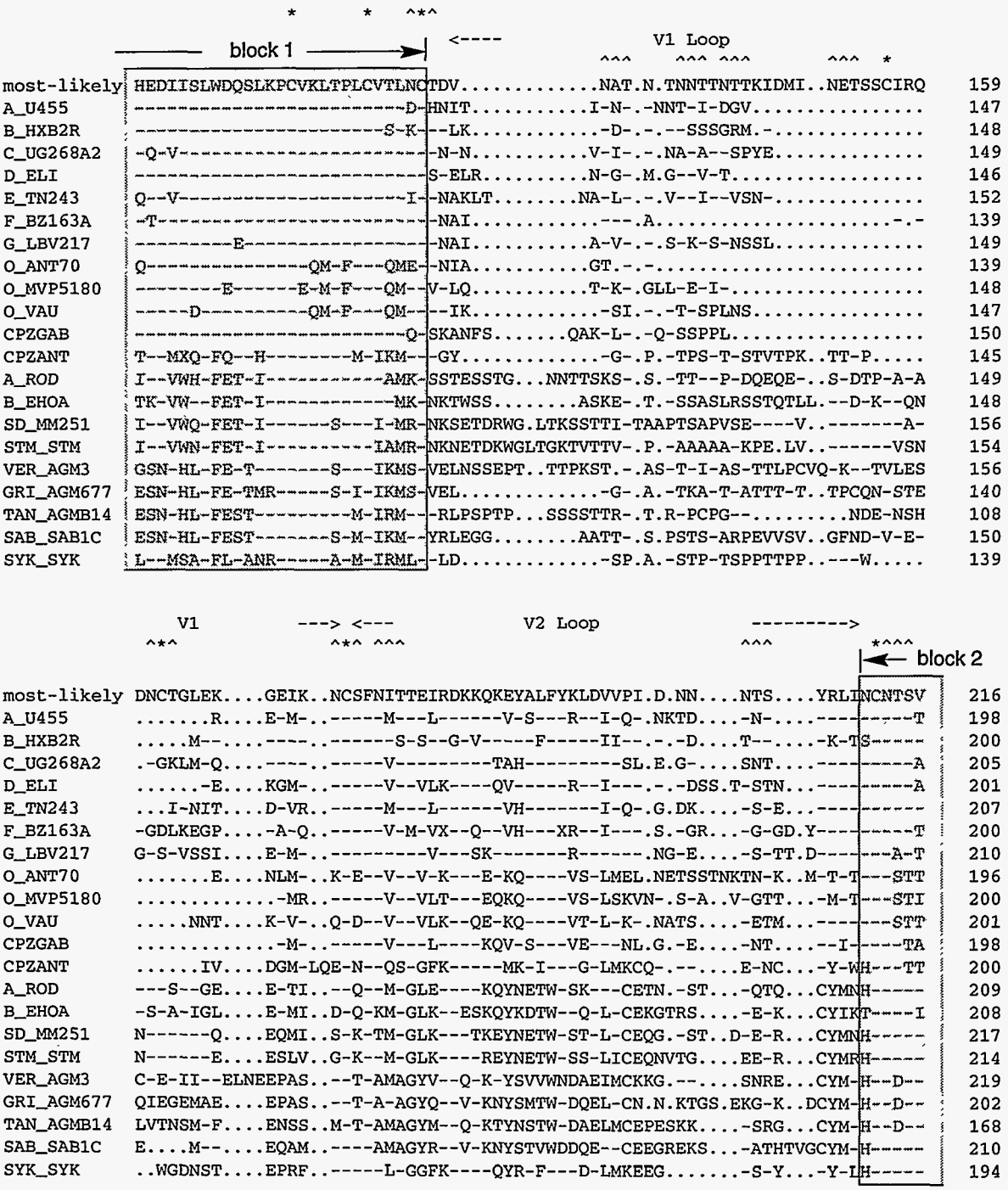
```
                                                                  env cds ->
                                                                   phos |    | phos
CONSENSUS-A   mtPL???eIcAIvGLiVALILAIVVWTIVgI.eyKkllkqr.......Kidrl?ikRIrERA.EDSgNES     57
CONSENSUS-B   -qs-...q-?---a-v--a-i--------f-?--r-i-R--......?-------d------.-------     56
CONSENSUS-D   -Q--...v-l---A-v----i--------f-.-crr-kr--.......---w-.-d------?-------     57
CONSENSUS-F   -S??...LAIS?TA------I-------?Y-.--R---R--........--N--.YE?--?--.-------     51
CONSENSUS-O   -H??...?LL-?I??SAL??INV??-?..F?..LR?Y-?-??QDR?E?E-LER.LR--?-IR.D--DY--     42
CONSENSUS-U   -Q--...T-T-----V--F-A-----S--Y-.--R-IR--K.......-----.LD------.-------     57
CONSENSUS-CPZ --??...?????L???????W?-CI???I????-??YK???....??????-?.??I?????.??????-     14

CONSENSUS-A   ?GDT?E.L?kL....VEM.GnydlgvdnNL$                                          78
CONSENSUS-B   e--qe-.-sa-????---?-H?apwdvdD--                                          79
CONSENSUS-D   E--rE-.-sa-....---.-HhAPwd?Ddm-                                          80
CONSENSUS-F   E--AE-.-A?-....G--.-PFIP-DI?---                                          73
CONSENSUS-O   N?EE-QEVM?-....??SH-F?NPM.FE??                                          59
CONSENSUS-U   D---E-.-ST-....M--.-YEYILDND---                                          81
CONSENSUS-CPZ -?EE--??-???????????FANP?.????DE                                         23
```

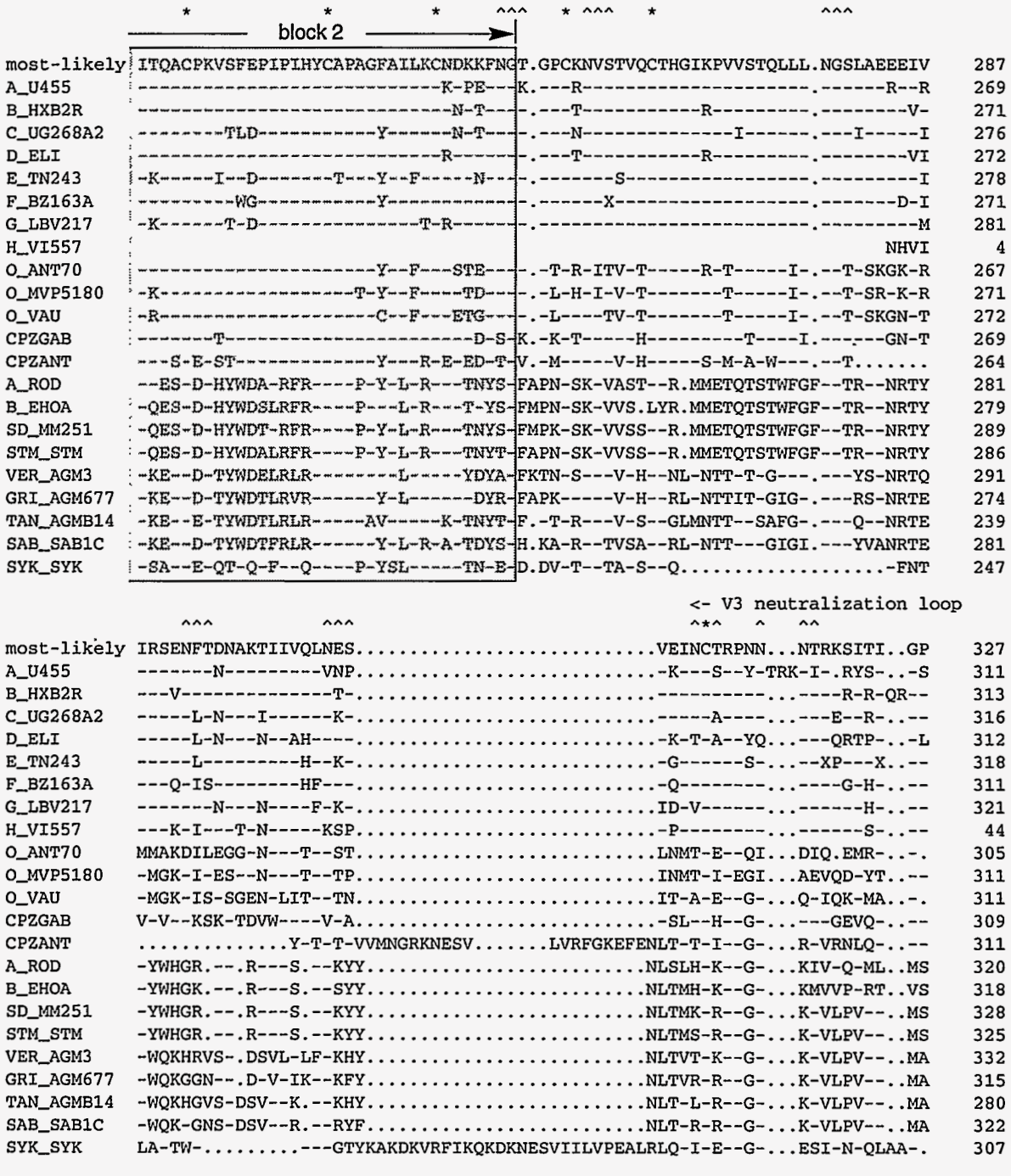
HMMER Sequences in the Env Alignment

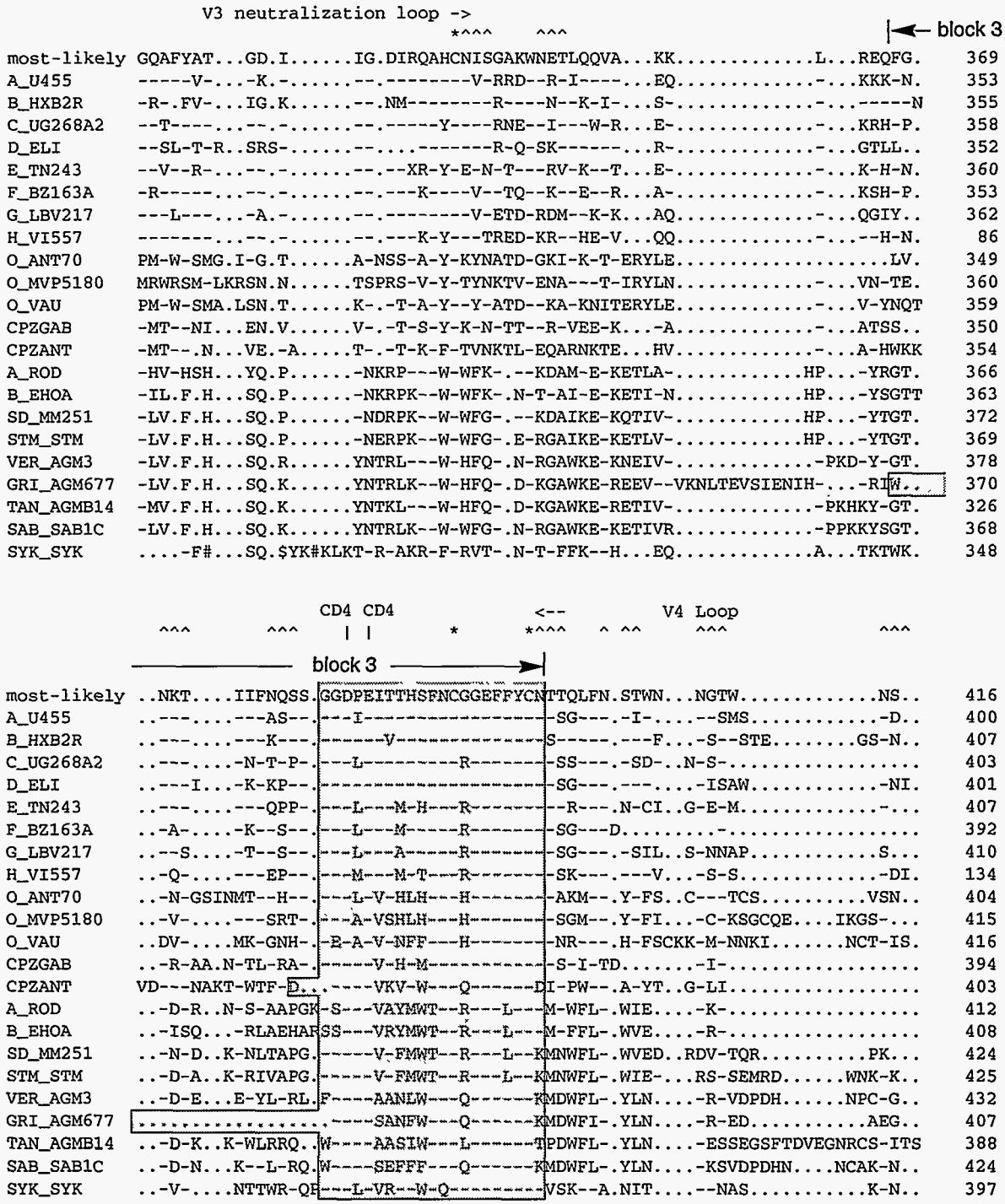
A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_HXB2R	HIVHXB2R	K03455	Starcich,B.	Science 227, 538 (1985)
C_UG268A2	HIVUG268A2	L22948	Louwagie,J.J.	JVI 69, 263 (1995)
D_ELI	HIVELI	K03454	Alizon,M.	Cell 46, 63 (1986)
E_TN2432	HIVTN2432	L03703	McCutchan,F.E.	ARHR 8, 1887 (1992)
F_BZ163A	HIV1BZ163A	L22085	Louwagie,J.J.	ARHR 10, 561 (1994)
G_LBV217	HIVLBV217	L11778	Louwagie,J.J.	AIDS 7, 769 (1993)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68, 1586 (1994)
O_MVP5180	HIVMVP5180	L20571	Gurtler,L.G.	JVI 68, 1581 (1994)
O_VAU	HIV1VAU	X80020	Charneau,P.	Virology 205, 247 (1994)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
A_ROD	HIV2ROD	M15390	Clavel,F.	Nature 324, 691 (1986)
B_EHOA	HIV2EHOA	U27200	Rey-Cuille,M.A.	Virology 202, 471 (1994)
SD_MM251	SIVMM251	M19499	Franchini,G.	Nature 328, 539 (1994)
STM_STM	SIVSTM	M83293	Novembre,F.J.	Virology 186, 783 (1992)
VER_AGM3	SIVAGM3	M30931	Baier,M.	Virology 176, 216 (1990)
GRI_AGM677	SIVAGM677	M66437	Fomsgaard,A.	Virology 182, 397 (1991)
SYK_SYK	SIVSYK	L06042	Hirsch,V.M.	JVI 67, 1517 (1993)

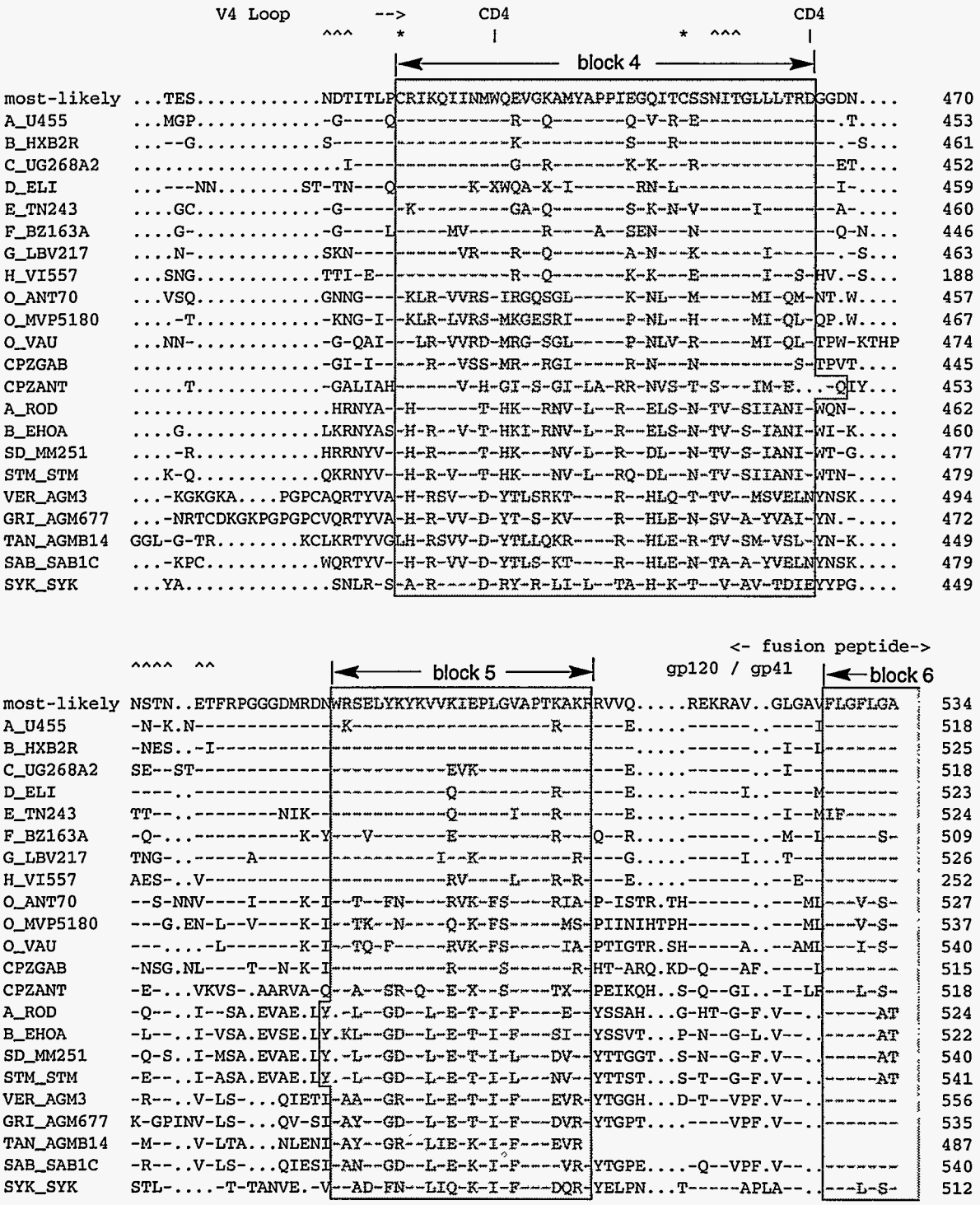
The following alignment and most-likely sequence were generated by the HMMER program as described in Part III. Approximately 400 HIV and SIV sequences contributed to the output; for simplicity, only representative subtypes and types are shown. Hybrid (i.e. recombinant) viral sequences will not affect the profile. The annotation was based upon information for HIV-1s and therefore the user should be cautious about its applicability to other PIV sequences. Cysteines are indicated by '*'. Potential N-linked glycosylation sites are indicated by '^^^'. Four highly conserved residues involved in CD4 interaction have been annotated. Eight conserved, gapless blocks (shaded) were independently derived from the representative sequences below (excluding the most-likely sequence) using the BLOCKMAKER and Motif programs; in most instances, the boundaries of these blocks are consistent with the HMM alignment. Block sizes are limited to 55 residues, hence contiguous blocks may occur. Five additional blocks over the C-terminal region of gp41 were separately determined, as variability and the existence of prematurely truncated (or unsequenced) representatives necessitated a separate determination. These five blocks should be given less weight. The Motif "Cobbler" sequence (involving just the first eight blocks) follows this alignment.

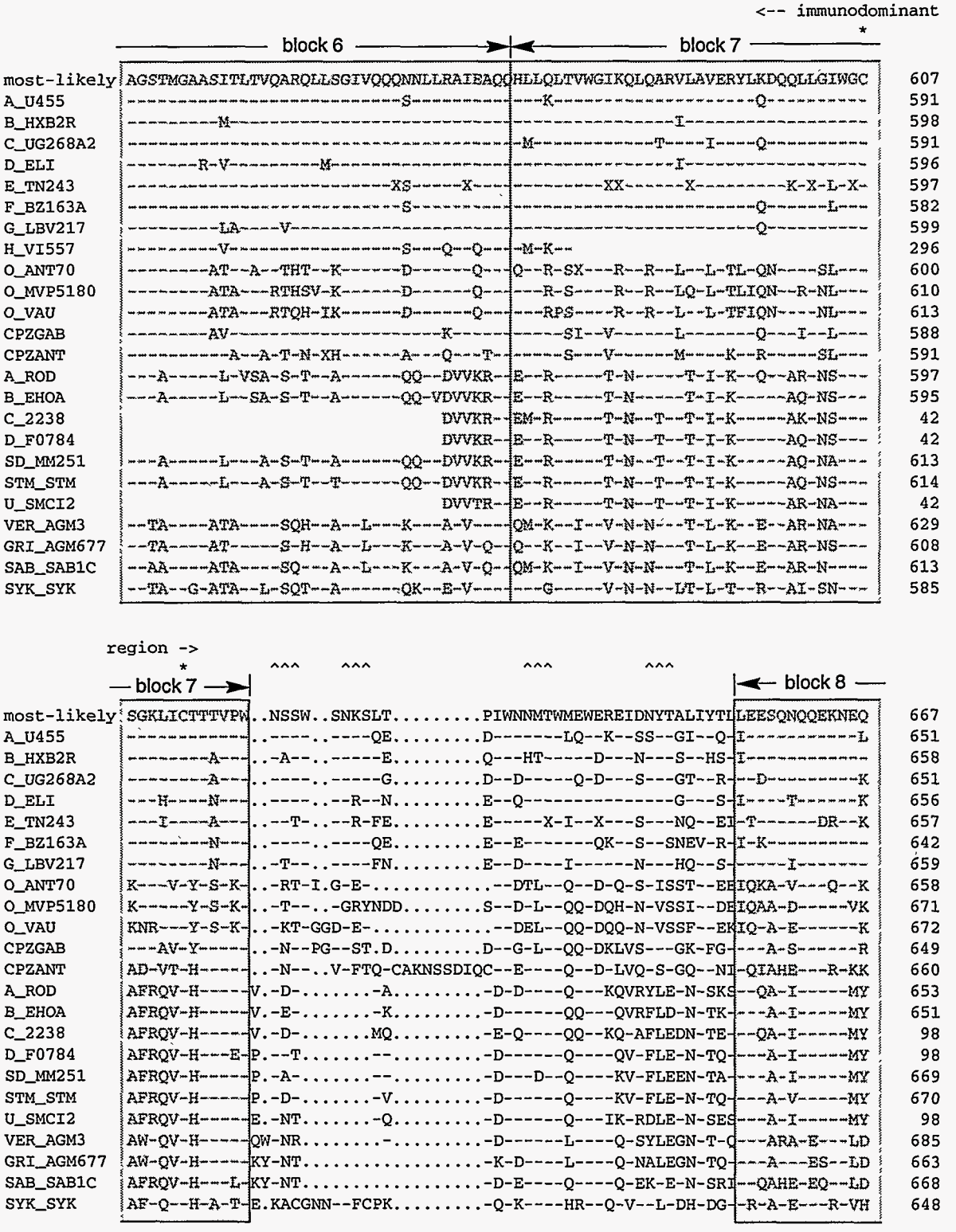
```
                                                    <- vpU
                                              signal peptide / gp120
                                                          *
most-likely MR.V....KGIRRNYQHLWRW..............GI....LLLGMLMI.CS.AAENLWVTVYYGVPVWKE     46
A_U455      --.-....M--Q---PC----..............-T....MI--LII-.-N.-.QQ-------------D     45
B_HXB2R     --.-....-E...K-------..............-WRWGTM--------.--.-T-K--------------     47
C_UG268A2   --.-....M--Q--C-QW-I-..............--....-GFWI---.-N.VMG---------------     46
D_ELI       --.A....R--E--C-NW-K-..............--....M---I--T.--.--D---------------     46
E_TN243     --.-....-ETQM-WPN--K-..............-T....-I--LVI-.--.-SD-------------RD     46
F_BZ163A    --.-....R-MQ--W---GK-..............-L....-F--I-L-.-N.-.---------------D     45
G_LBV217    --.A....---L--W----K-..............--....-I--LVI-.-N.-SG-----------A-ED     46
O_ANT70     -I.-TMKAME.K--.KK--..................TL....Y-AMA-ITP-L.SLRQ-YA---A-----ED     47
O_MVP5180   -T.-T..M-VMKK-NRKS-..................SL....YIAMA-L-P-LSYSKQ-YA---S-----E-     48
O_VAU       -T.AIMKAM-.K--.RK-GI-..............CL....I-ALIIPC.L-.C.NQ-YA---S-----ED     47
CPZGAB      -K.-....MEKKKR.......DWNS.........LSIIT....IITII-LTP-L.TS-.------------HD     46
CPZANT      --.K...........PIHII-..............-L....A--IQFIE.KG.TN-.DY---F------RN     38
A_ROD       -M.N...........-L-I..................A-....--ASACLV.Y..C.T.QY---F----T--N     34
B_EHOA      -AH-.........-NYL-V.....................T-LLIS-.YG.YMGKNF---F--I-A--N     37
SD_MM251    -G.C.......LG-.-L-I..................A-....---SVYG-.Y..C.T.QY---F----A-RN     37
STM_STM     -A.C.......PG-.-L-I..................A-....---SACLT.Y..C.T.QY---F----A-RN     37
VER_AGM3    -K.L...........TL-I..................--....--I-IGVV.LN.T.RQQ----F-------N     36
GRI_AGM677  -G.R...........L-I..................K....I-IIAIG-.SI.GIG--Y---F--I----N     35
TAN_AGMB14  XQ.A....FCMTP-..........................................................      9
SAB_SAB1C   -K.L.................................LT....V--WLSGC.W-LVWLVQY---F--I----N     34
SYK_SYK     -A.A.................FRTYIVCLFSLISLGF.................MEKQQY---F--I-H-ED     37

                  *                   *                ^^^
most-likely ATTTLFCASDAKAYDTEVHNVWATHACVP..TDP..NPQ.EIVLENVT.ENFNMW..K..NN.M....V.EQM    103
A_U455      -V-------------A-------------.---.---.--D-V---.-E----..-..--.-....-.D--    102
B_HXB2R     ----------------------------.---.---.-V--V---.------..-..-D.-....-.---    104
C_UG268A2   --------------E-------------.---.---.--------.-S----..-..-D.-....-.D--    103
D_ELI       ------------S-E--A--I--------.---.---.--A-----.------..-..--.-....-.---    103
E_TN243     -D-----------HE-------------.---.---.--Y-----.-------..N..--.-....-.---    103
F_BZ163A    ------------S-ER-------------.---.---.-V------.---D--..-..-S.-....-.---    102
G_LBV217    -D------------S--S-----------.---.---.--S-N---.------..-..--.-....-.---    103
O_ANT70     --PV-------NLTS--K--I--SQ----..---..T-Y.-YP-H---.DD--I-..-..-Y.-....-.---    104
O_MVP5180   -APV-------NLTS--Q--I--SQ----..---..--H.-FP-G---.D--DI-..-..-Y.-....-.D--    105
O_VAU       -KP--------NLTS--Q--I---Q----..---..S-N.-YE-K---.GK--I-..-..-Y.I....-.D--    104
CPZGAB      -DPV--------HS--A--I---Q----..---..S--.-VF-P--I.-S----..-..--.-....-.D--    103
CPZANT      --P-----TN-SMTS---------TS---..I--..D-I.VVR-.-TS.VW--AY..-..-Y.-....-.-S-     94
A_ROD       --IP----TR.NR.....D.T-G-IQ-.L..P-N..DDYQ--T-.---.-A-DA-..N..-..T....-T--A     83
B_EHOA      -SIP----TR.NR.....D.T-G-VQ-.L..P-N..DDYT--Q-.-I-.-A-DA-..D..-..T....-TD-A     86
SD_MM251    --IP----TK.NR.....D.T-G-TQ-.L..P-N..GDYS-LA-.---.-S-DA-..E..-..T....-T--A     86
STM_STM     --IP----TK.NR.....D.T-G-TQ-.L..P-N..GDYS-LAI.---.-A-DA-..D..-..T....-T--A     86
VER_AGM3    SSVQA--MTPTTR.......L---TNS.I..P-D..HDYT-VP-.-I-.-P-EA-ADR..-P.L....-.A-A     87
GRI_AGM677  S-VQA--MTP.NT......-M---TN-I-..D-H..DNT.-VP-.-I-.-A-EA-..D..-P.L....-.K-A     84
TAN_AGMB14  .................T-L---TN-I-..D-H..-YT.-VQ-.--S.-K-EA-..-DR-P.L....-.A-A     49
SAB_SAB1C   SSVQA--KTP.NT......-L--STN-I-..D-EPEGTIA-VPIP-I-.-K-DA-..-..-...RNPL-.G-A     89
SYK_SYK     -YAP---TTSH-G.......G---KN---SA......D-I-VRV.-I-G-Y-PA-..N..SSH-....IRQ-I     88
```

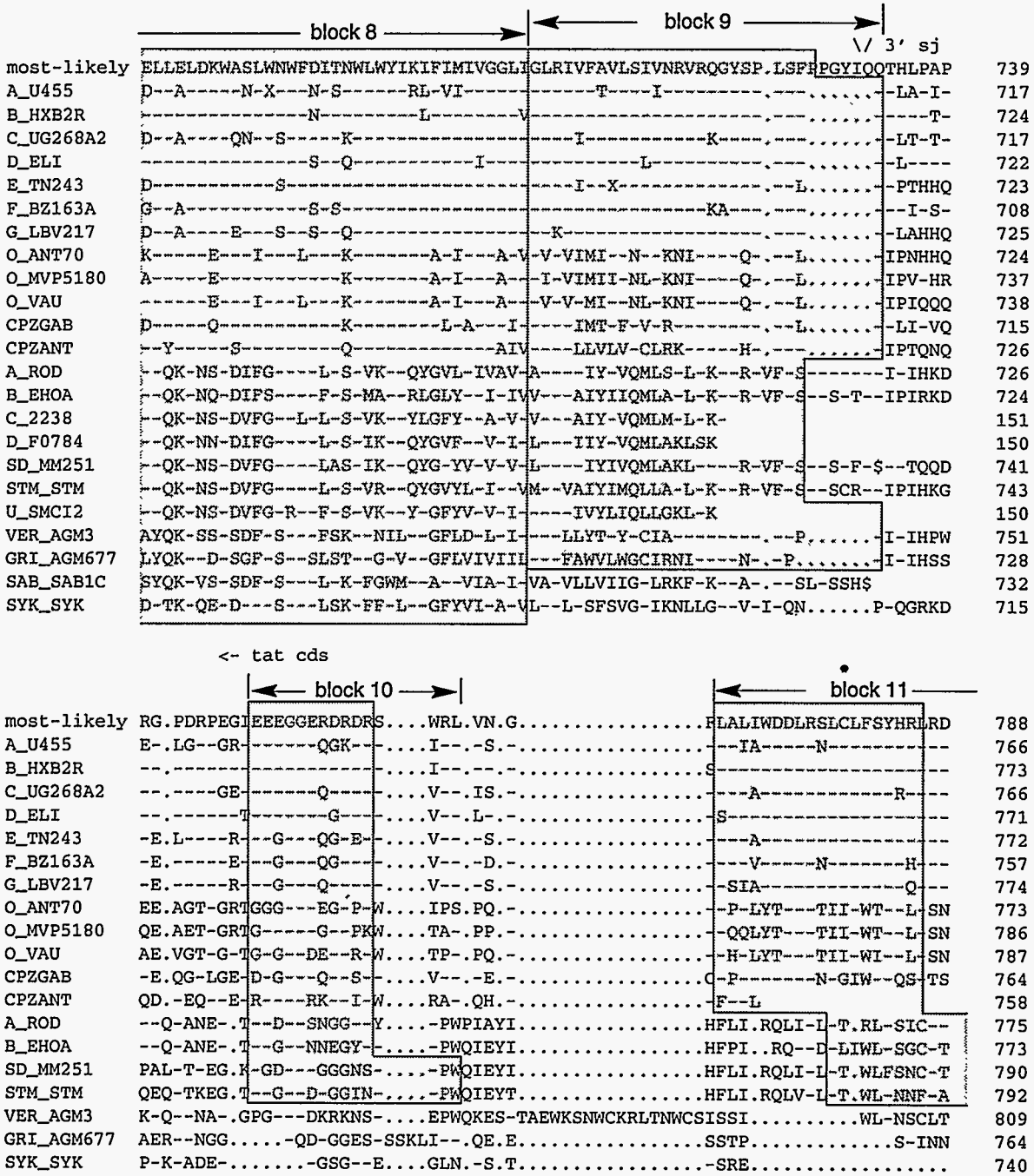

block 1
V1 Loop
V1
V2 Loop
block 2

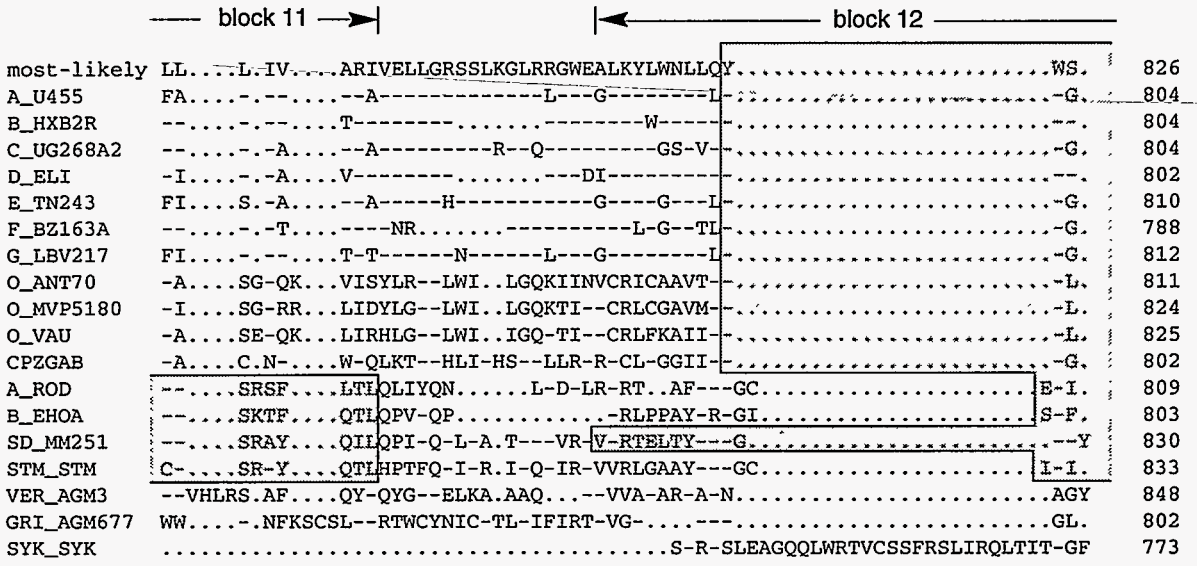

block 2
<- V3 neutralization loop

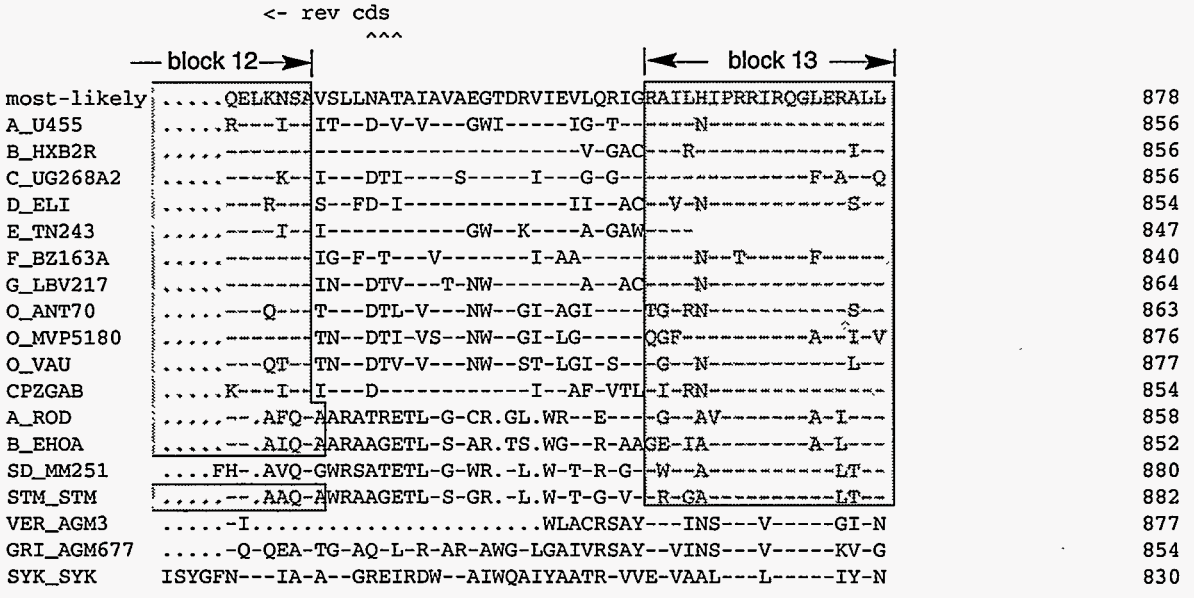

V3 neutralization loop ->
block 3
CD4 CD4
V4 Loop
block 3
most-likely
A_U455
B_HXB2R
C_UG268A2
D_ELI
E_TN243
F_BZ163A
G_LBV217
H_VI557
O_ANT70
O_MVP5180
O_VAU
CPZGAB
CPZANT
A_ROD
B_EHOA
SD_MM251
STM_STM
VER_AGM3
GRI_AGM677
TAN_AGMB14
SAB_SAB1C
SYK_SYK


V4 Loop
-->
CD4
CD4
block 4
block 5
gp120 / gp41
<- fusion peptide->
block 6
most-likely
A_U455
B_HXB2R
C_UG268A2
D_ELI
E_TN243
F_BZ163A
G_LBV217
H_VI557
O_ANT70
O_MVP5180
O_VAU
CPZGAB
CPZANT
A_ROD
B_EHOA
SD_MM251
STM_STM
VER_AGM3
GRI_AGM677
TAN_AGMB14
SAB_SAB1C
SYK_SYK


<-- immunodominant
*
block 6
block 7
region ->
*
^^^
^^^
^^^
^^^
block 7
block 8


block 8
block 9
\/ 3' sj
most-likely ELLELDKWASLWNWFDITNWLWYIKIFIMIVGGLIGLRIVFAVLSIVNRVRQGYSP.LSFHPGYIQQTHLPAP 739
A_U455 D--A-----N-X---N-S------RL-VI-----------T----I---------.---.....--LA-I- 717
B_HXB2R ----------N---------L---------V-------------------.---.....-----T- 724
C_UG268A2 D--A----QN--S-----K-----------------I----------K-----.---.....-LT-T- 717
D_ELI ----------S--Q----------I-----------L----------.---.....-L---- 722
E_TN243 D----------S-------------------I--X----------.--L.....-PTHHQ 723
F_BZ163A G--A----------S-S-----------------KA---.---.....--I-S- 708
G_LBV217 D--A----E---S--S--Q----------K----------.---.....-LAHHQ 725
O_ANT70 K-----E---I---L---K-------A-I---A-V-V-VIMI--N--KNI----Q-.--L.....-IPNHHQ 724
O_MVP5180 A-----E----------K-------A-I---A--I-VIMII-NL-KNI----Q-.--L.....-IPV-HR 737
O_VAU ------E---I---L---K-------A-I---A--V-V-MI--NL-KNI----Q-.--L.....-IPIQQQ 738
CPZGAB D-----Q----------K-------L-A---I-----IMT-F-V-R-------.--L.....-LI-VQ 715
CPZANT --Y-----S---------Q----------AIV----LLVLV-CLRK-----H-.---.....-IPTQNQ 726
A_ROD --QK-NS-DIFG----L-S-VK--QYGVL-IVAV-A----IY-VQMLS-L-K--R-VF-S-------I-IHKD 726
B_EHOA --QK-NQ-DIFS----F-S-MA--RLGLY--I-IVV---AIYIIQMLA-L-K--R-VF-S--S-T--IPIRKD 724
C_2238 --QK-NS-DVFG--L-L-S-VK--YLGFY--A-V-V---AIY-VQMLM-L-K- 151
D_F0784 --QK-NN-DIFG----L-S-IK--QYGVF--V-I-L---IIY-VQMLAKLSK 150
SD_MM251 --QK-NS-DVFG----LAS-IK--QYG-YV-V-V-L----IYIVQMLAKL----R-VF-S--S-F-$--TQQD 741
STM_STM --QK-NS-DVFG----L-S-VR--QYGVYL-I--VM--VAIYIMQLLA-L-K--R-VF-S--SCR--IPIHKG 743
U_SMCI2 --QK-NS-DVFG-R--F-S-VK--Y-GFYV-V-I----IVYLIQLLGKL-K 150
VER_AGM3 AYQK-SS-SDF-S---FSK--NIL--GFLD-L-I----LLYT-Y-CIA--------.--P.....-I-IHPW 751
GRI_AGM677 LYQK--D-SGF-S--SLST--G-V--GFLVIVIII---FAWVLWGCIRNI----N-.-P.....-I-IHSS 728
SAB_SAB1C SYQK-VS-SDF-S---L-K-FGWM--A--VIA-I-VA-VLLVIIG-LRKF-K--A-.--SL-SSH$ 732
SYK_SYK D-TK-QE-D---S---LSK-FF-L--GFYVI-A-VL--L-SFSVG-IKNLLG--V-I-QN.....P-QGRKD 715
<- tat cds
block 10
block 11
most-likely RG.PDRPEGIEEEGGERDRDRS....WRL.VN.G.....RLALIWDDLRSLCLFSYHRLRD 788
A_U455 E-.LG--GR-------QGK--....I--.-S.-.....--IA-----N---------- 766
B_HXB2R --.-------------....I--.--.-.....S---------- 773
C_UG268A2 --.----GE-------Q-----....V--.IS.-.....---A----------R---- 766
D_ELI --.------T-------G---....V--.L-.-.....S---------- 771
E_TN243 -E.L----R---G---QG-E-....V--.-S.-.....---A---------- 772
F_BZ163A -E.-----E---G---QG---....V--.-D.-.....---V-----N-------H--- 757
G_LBV217 -E.-----R---G---Q-----....V--.-S.-.....-SIA----------Q--- 774
O_ANT70 EE.AGT-GRTGGG---EG-P-W....IPS.PQ.-.....-P-LYT---TII-WT--L-SN 773
O_MVP5180 QE.AET-GRTG-----G--PKW....TA-.PP.-.....-QQLYT---TII-WT--L-SN 786
O_VAU AE.VGT-G-TG-G--DE--R-W....TP-.PQ.-.....-H-LYT---TII-WI--L-SN 787
CPZGAB -E.QG-LGE-D-G---Q--S-....V--.-E.-.....C-P-------N-GIW--QS-TS 764
CPZANT QD.-EQ--E-R----RK--I-W....RA-.QH.-.....F--L 758
A_ROD --Q-ANE-.T--D--SNGG--Y....-PWPIAYI.....HFLI.RQLI-L-T.RL-SIC-- 775
B_EHOA --Q-ANE-.T--G--NNEGY-....-PWQIEYI.....HFPI..RQ--D-LIWL-SGC-T 773
SD_MM251 PAL-T-EG.K-GD---GGGNS-....-PWQIEYI.....HFLI.RQLI-L-T.WLFSNC-T 790
STM_STM QEQ-TKEG.T--G--D-GGIN-....-PWQIEYT.....HFLI.RQLV-L-T.WL-NNF-A 792
VER_AGM3 K-Q--NA-.GPG---DKRKNS-....EPWQKES-TAEWKSNWCKRLTNWCSISSI.....WL-NSCLT 809
GRI_AGM677 AER--NGG.....-QD-GGES-SSKLI--.QE.E.....SSTP.....S-INN 764
SYK_SYK P-K-ADE-.....-GSG--E....GLN.-S.T.....-SRE..... 740





COBBLER sequence from MOTIF

>Env C_UG268A2, with embedded consensus blocks

mrvmgiqrncqqwwiwgilgfwilmicnvmgnlWVTVYYGVPVWKDATTTLFCATdakay
etevhnvwathacvptdpnpqeivlenvtesfnmWKNTMVEQMHEDIISLWEQSLKPCVK
LTPLCIKMNCtnvnvnitnnanatnspyengklmeqgeiknсsfnvtteirdkkqtahal
fykldvvslegnsntyrliHCNTSVITQACPKVYWDPIPIHYCAPAGYAILKCNDTNFNG
tgpcnnvstvqcthgikpvistqlllngsiaeeeiiirsenltnnakiiivqlnksvein
carpnnntresirigpgqtfyatgdiigdirqaycnisrnewnitlqwvreklkrhfpnk
tinftqpsGGDPEVTHHWFNCRGEFFYCNTSWFFNssdnnnstiitLPCRIRQIVNDWHK
VGKAIYAPPIEGNITCNSNITGLLLTLDggetsetnstetfrpgGGDMKDIWRSELYKYK
VVEIKPIGVAPTKVKRrvverekravgigavflGFLGAAGSTMGAASTTLTVQSRQLLSG
IVQQQQNLLRAIEAQQHLLRLTVWGVKQlQARVLAIEKYLQDQQLLNLWGCAGKLICHTT
VPWnsswsnkslgdiwdnmTWQEWERQINNYTGNIYQLLEEAQNQQEKNEKELLKLDKWS
SFWNWFDITKWLWYikifimivggliglriifavlsivnrvrkgysplsfqtltptprgp
drpgeieeeggeqdrdrsvrlisgflalawddlrslclfsyrrlrdllliaaravellgr
sslrglqrgwealkylgslvqywgqelkksaislldtiaiavsegtdriievgqgigrai
lhiprrirqgfeaalq

```
                                                  <- vpU cds
                                          signal peptide / gp120
                                                     *                                  *
CONSENSUS-A     Mrvmgiq?nyq?l.wr??....??W.gtmilg??iIc.na??e.?lWVtVyYGVPVWkdaeT..TLfc      47
CONSENSUS-B   ??---k--rk---h-?----????---.---l--mlm--.s----.--------------e-t-..----      51
CONSENSUS-C     ---?--?r-?-qw.-i...???---.-ilgfwmlm--.-v--g.n-------------e-k-??----      49
CONSENSUS-D     ---r?-er---h-.----????---.---L--mLM--.sv.a??-------------E-t-..----      50
CONSENSUS-E     ---Ket-m-wpn-.-k--....---.--l---lv---?s-.Sd.N------------r--d-..----      53
CONSENSUS-F     ---R-M-R-W-H-.GK........-.-LLF--iL---.--.--.n-------------e-T-..----      52
CONSENSUS-G     -?-k---r-W-H-.-k........-.--L---LV---.s-.sn.n------------E--D-..----      52
CONSENSUS-H                                                                  ..          0
CONSENSUS-O     -?-tMk?MeKrN?.Kl........-il?l?mALi-P-.LS.-n?q-YA---?-----E--?P..v---      46
CONSENSUS-U     ----e--R-cp-?.-?........-.-il---m-m--.--.-??ns----------------..----      47
CONSENSUS-CPZ   -??????-???-?.??--?.....??????-?--???.?T.--.-??---?------??-?P..?---      18
                                *             ^^^                                 *
CONSENSUS-A   ASdAkAydtE?HNVW?aTHaCVPTDPnPqEi?le.NVTE?FnmwkNnMVeQmheDiiSLWD.qSLkPCvk     111
CONSENSUS-B   ----------v----.--------------vv-??----n---------------------.--------     117
CONSENSUS-C   -------e--v----.--------------?v--.----n------d--d-----------.---------     115
CONSENSUS-D   -----s-k?-a--i-.------------------.----N---------------------.---------     115
CONSENSUS-E   ------He--v----.--------------h--.----n-----------q--v-----?---------     120
CONSENSUS-F   -----S-Ek-v----.--------------Vv--.----n-d----------T--------.---------     119
CONSENSUS-G   -------s--s----.------------------.----n----------------------.E-------     118
CONSENSUS-O   ----NLTS--q--I-.-sQ-------t-?-yp-k.---dn--I---Y--?--?-------?.-------q     109
CONSENSUS-U   -----s-??------.----------s---?---.----n------?--------------.---------     108
CONSENSUS-CPZ -??-???S-----?-.--??---?--?-??V--?.-???---??--?--??-?--???-??.--?-----      54
                  *   ^* ^          ^^^          ^^^       ^^^
CONSENSUS-A   LtPLcvtl?c.???????????n?t???????????n?t???????n?????...????????........     125
CONSENSUS-B   --------n-.td---------?-?--------------------------???--------??...???     132
CONSENSUS-C   --------n-.-------...----------.--------------t---?...-------........     131
CONSENSUS-D   --------n-.t------------------....----------?-----?....------????????     130
CONSENSUS-E   --------n-.tna---------l--------nv--i-nvsniig-it..................???     151
CONSENSUS-F   ----------?--a--....---------------?-?-q--.........................?tl     134
CONSENSUS-G   --------n-.-----.....-v--t--.....?-?........NcT--??n--nNstv-...?????     141
CONSENSUS-O   M-F---QMn-.t------....-i-------....-----.....------...??????     125
CONSENSUS-U   --------n-.t-----....-------..-------?--.............------.........     122
CONSENSUS-CPZ ---?-???--.---------.............?-?-----.....-------P????..?      60
                       ^*^ ^^^                    ^^^     ^^^        ^^^
CONSENSUS-A   .m?..?e.....ikncsfnmttelrdkkqkvyslfyrldvvqi????????n??????....n???????     160
CONSENSUS-B   ?-ek?g-?????-------i--si---v--e-a---k----p-d-------?------????--------     170
CONSENSUS-C   ??-??--..???-------?---?--------a------i-pl---------------?...--s-----     164
CONSENSUS-D   ??-?.-g.....m------i--?v----kq-ha---k---------------------...?--t--.--     164
CONSENSUS-E   ??-..d-.....Vr-----------------hA---k--i---------.---s----?...?--...--     186
CONSENSUS-F   keep.ga.....-Q---------v---Ql--Ha------I-p-s---...--ns----...??--..---     175
CONSENSUS-G   ??-??e-?.???m------i---i----ktE-A---k----p-n---..--?ss-.......?--..---     177
CONSENSUS-O   ??-??n-??...?-q-e--v--V-k---E-KQA---Vs-L-k?n-ts--..--T---.............     161
CONSENSUS-U   ..-??-?.....?------?---?----kte?a---k----p-nd--...--------....?--.....     152
CONSENSUS-CPZ ??-??-?.....???-?--??-???---??????--??-????--......--.........--T.....      73
                    *^^^     *            *          * ^^^ ^^^  * ^^^   *
CONSENSUS-A   ??yrlincntsaitqacpkvsfepipihycapagfailkc?dk?fngtgpcknvstvqcthg?ikpvvst     225
CONSENSUS-B   ------s----v----------------------------n--k-------t----------.-r-----     237
CONSENSUS-C   ----------------------d-----------y-----nn-t-------?----------.-------     230
CONSENSUS-D   --------------------t------------------n--k------------------.-r-----     231
CONSENSUS-E   -----------V-k-----i--D-------t---y-----N--n-----------S------.-------     253
CONSENSUS-F   -----------T---------Wd-----------Y-----N--k------------------.-------     242
CONSENSUS-G   sd-------v-t-K------?-d----------------r---------------------.-------     244
CONSENSUS-O   .m-t-?---STt-k--------------------y--F--N?Te-----?---itV-T----.---T---     226
CONSENSUS-U   .------------?-------------------------n--K------------------.-------     218
CONSENSUS-CPZ ---????---T?---?-?-??-------------?---?---?D-?-?-?-?---?-H----.-?-?-?-     119
                                                        <- V3 neutralization loop
                    ^^^           ^^^          ^^^   ^*^    ^^^
CONSENSUS-A   qlll.ngslae??ev?irSenitnNaktiiVql??pV?InCtRp.nnntr.ks???vri???gpGq??af     278
CONSENSUS-B   ----.------e.--v-----f-d---------nes-e------?-----.----.ih-------r..--     295
CONSENSUS-C   ----.--------.-ii-----l---?-----h-n-s-e-v----.-----.--...i--...-----..t-     285
CONSENSUS-D   ----.------E.-iI-----l-----?-----nes--------?y?---.qr..-tp-...-?--..-l     285
CONSENSUS-E   ----.------e.-Ii-----L---------h-NKs-e------.s----.t-...it-...-----..v-     311
CONSENSUS-F   ----.------e.dii---q--sd---------Nes--------.-----.--...I?-...---r..--     298
```

HIV1 ENV CONSENSUS

```
CONSENSUS-G   ---?------e.-I------?-d---v-----n-sie-?----.-----.--...i?f...----..--   297
CONSENSUS-H        .      -.?-----?---D-T-N-----K----------.-----.?---.I?-...-?--..?-    38
CONSENSUS-O   --I-.--T-Skg.kIr-Mgk--?dsg-N---T-N-?i-mt-e--.g-?-v.Qe...i?-...--.m..-W   277
CONSENSUS-U   ----.------E.-?i-------d---?-----ne?--------??----?--...?--...-----..?-   269
CONSENSUS-CPZ ?--?.--????--??-?????K?????V?????-E??-??-?--.G-?-?.??...?Q-...---M..T-   140
          V3 neutralization loop ->
                                *^^^    ^^^                    ^^^          ^ ^^
CONSENSUS-A   ya.tgdi.....iG.dirqAhCnvsr?eWn?tlq?v....a?qLr?..yf???nkt....??iiF?n.ss   319
CONSENSUS-B   -t.--?-???..--.--------i--ak--n--kqi....v-k--e??q-.-----?...---v-nq?--   340
CONSENSUS-C   --.----.....--.--------i-?-k--------....?kk-ae..h-p-----.....-?--?.--   325
CONSENSUS-D   -t?-?r????..?-?--------i-?a?--k---q-....--k-gd?.ll..----.....t---kp.--   327
CONSENSUS-E   -r.----.....--.---k-y-EINgTk--e?-kq-....tek-ke..H--.n---.....---qP?p-   358
CONSENSUS-F   --.--?-.....--.---k------gtq-----e--....?ak-ks..h--.--?-.....--k-ns.--   340
CONSENSUS-G   --.--?-.....--.----------?-?-?em--n-....?-?-?-..?------?...?---t-ns?-a   332
CONSENSUS-H   ?-.?-?-.....--.---?-?--I??-?-?-?-?--....?---?-..H?..--??.....--?--P.--    62
CONSENSUS-O   -S.M-.l???n?k???s-?-Y-?YnaTd-?ka-kqt....-eRYLe..Lv...-?-????vtm?-n?s-?   318
CONSENSUS-U   --.--??.....--.--------i-?t?------q-.....--?---..?--.--?-.....-----??--   304
CONSENSUS-CPZ -N.?E??.....?-.-T-?-?-??N?T?-?-???-??????-???-..??...-?-...?A-???-?.??   155
          CD4
            | | CD4   *       *^^^    ^^^           ^^^          ^^^          ^^^
CONSENSUS-A   GGDlEitthsfnCggef?FYCnts?lF.nstw???????....n?t??????????n?t????..????n   353
CONSENSUS-B   ---p--vm---------.-----tq--.----------????-------------------??----?   374
CONSENSUS-C   ------------r---.---------?---y-------????----------------?----??----?   357
CONSENSUS-D   ---p-----------.---------?-?---------????--?-----------------??----?   358
CONSENSUS-E   -------m-h---r---.-----t---.-n-cig-........-e-m-----gc..-g-.........-?   398
CONSENSUS-F   -------m-----r---.---------.-d-?------.....-------.....?-?---...----?   370
CONSENSUS-G   ------------r---.------g--?-?s?---.......?----------.---?----??----?   363
CONSENSUS-H   ---?--??-?---?---.-----?K--.--------.....--?---.N---..-G-...........    91
CONSENSUS-O   ---?-v-hlh---H---.-------m-.-y-Fsc-----????----------------n--.....--   354
CONSENSUS-U   ---------t-------.---------.-??----.....?-?--------.---?--....----?   332
CONSENSUS-CPZ ?--P-V??-?---?---.---???-?-.????---.....?I-.....-                      174
                  *      CD4 |                  *  ^^^      | CD4    ^^^
CONSENSUS-A   ??It..lq.CrI.kqivnm.wQrvgq.amYapPI?.g?irc?snITGllLTrDgg??....nns??...?   399
CONSENSUS-B   ----??-p.---.---i--.--e--k.---------.-q---s---------------????-??--?.?-   420
CONSENSUS-C   ----..-p.---.---i--.--e--r?---------.-n-t-k---------------????-?t--???-   404
CONSENSUS-D   -t--..-p.---.---i--?--g--k.--------e.---?-s---------------?...-?--...?-   405
CONSENSUS-E   ----..-P.-k-?---i--.--ga--.--------s.-r-n-V-----I-------a-??..--t-.....   448
CONSENSUS-F   -t--..-p?---.------.--e--r.----a--a.-n-t-n---------------??.?-?t--?...   419
CONSENSUS-G   ----..-p.---.----r-.------.-------A?-n-t-----------------?...-?t-.....   410
CONSENSUS-H   ---?..-?.---.---??-.---?--.---------.-K-T--------I--S-HV......-?-AE....   134
CONSENSUS-O   gT??..iP.--L.r-v-Rs.-m-G-S.gl-------.-NLt-------MI-Ql-?.pW....-s---....   401
CONSENSUS-U   ----..-P.---.----r-.--?---.---------.---?---------------n-?...-??--....   375
CONSENSUS-CPZ G-?I..??.---.?---??.-??-??.GI-??-?R.-N??---?---??-?????--....-??......   198
                 ^^^                                                  gp120 / gp41
CONSENSUS-A   ????netFrPgGgdmrdNWrsELYkYKvVkiePlGvaPtr.akrRvV....eREKRA??vg.lGavflgf   457
CONSENSUS-B   ----t-i-------------------------------k.------....q------..--?i--m----   478
CONSENSUS-C   ----?------------------------e-k-------?.------??..------.---?i-------   460
CONSENSUS-D   ----------------------------r----------.------....-------..I-.---m----   463
CONSENSUS-E   ...----------NiK-------------Q-----i----.------....------..--.I--Mif--   506
CONSENSUS-F   ..--?--------n-k-------------e----------.---q--....k---------.?--l----   475
CONSENSUS-G   ..----------------------------k--------.-R----....g-----..--.--------   468
CONSENSUS-H   ...-?V----------------------??----?---?.-?----....?-----..--.-?------   184
CONSENSUS-O   .----?-l--?----k-I--T--f-----rvK-FS----ki-RP?Igt?t?H-----..--.--ML---v   459
CONSENSUS-U   ..--?-----------------------------------.------....------.---.m--?----   431
CONSENSUS-CPZ ...-?????-??????-?--?--??-?--?-?--S----???R???????.?-Q--?.-?-.?-??---?   225
CONSENSUS-A   lGaAGstmgaasitltvqarqllsgivqqqsnllraieaqqhllkltVWGIKQLQARvLAvErYLrD.QQ   526
CONSENSUS-B   ------------?----------------n-------------q--------------------k-?--   546
CONSENSUS-C   ----------------------------------------m-Q----------t----i----k-.--   529
CONSENSUS-D   ------------?----------------N-------------Q--------------------k-.--   531
CONSENSUS-E   ----------------------------------------Q--------------------K-.-K   575
CONSENSUS-F   ----------------------------n-------------Q--------------------k-.--   544
CONSENSUS-G   ------------------V-----------------------Q--------------------?-.--   536
CONSENSUS-H   ------------?----------------------?--?----?----                       227
CONSENSUS-O   -S---------ATa----tht-?K------D------Q---?--R-S----R--R--L--L-TliQN.--   526
CONSENSUS-U   ----------------------------?-----------?-Q-----------?-------??.--   495
CONSENSUS-CPZ -?---------???--?-?-?-??------?---?---?-----Q-S?--V------?----?--?-.--   278
```

```
                  *     *      ^^^            ^^^      ^^^         ^^^
CONSENSUS-A   LLGIWGCSGKlICtTnVPWNsSW.........S.Nks??dIWdnMTWlqWdKEisnYT?iIY?.LiEesq    581
CONSENSUS-B   ---------------a----a--.........-?---l-?--?----me-er--d----l--t.------    601
CONSENSUS-C   ---------------a-------.........-?---q---------m---r-------t--r?-l-d--    586
CONSENSUS-D   ----------h-----------.........-.-r-L-e--?----mE-ER--d---Gl--s.------    588
CONSENSUS-E   f--L------I----A-----t-.........-.-r-fEE--n----iE-eR------nq--e.ILT---    634
CONSENSUS-F   ---L-------------------.........-.---qEe--g----ME-e------SnE--R.---?--    602
CONSENSUS-G   ----------------------t--.........-.---fnE-------Ie-eR--N----q--n.-l----    594
CONSENSUS-O   --nL---K-?---Y-S-k--kt-?g.......????dnes----L--Q?--qq-n-?SSt---.e-QkA-    579
CONSENSUS-U   ---?---?--?--?-?-------.........-.-----??-?-?--ME-?R--??---V---.-?-?--    536
CONSENSUS-CPZ ?-?L---??-??-?-T----N--??????????.???--?--?-?--Q?--?LV?-?-G?-?-.?L??A?    310
                                                                              \/ 3'sj
CONSENSUS-A   nqQEkNEqdLLaLDkWanLwnWFdIsnWLWYIriFimIVGGLIGLRIvfaVlsiInRVRqGYSPlSFQtl    651
CONSENSUS-B   --------e--e-----s-----?-t------k---------v----------v--------------?    669
CONSENSUS-C   -------k--------q--------t?-----k--------------i------v---------------    655
CONSENSUS-D   ?-------e--?-----S-----s-T?-----k--------------------lV---------------    655
CONSENSUS-E   ---DR--K---e-----S-------T------K--------------i------V--------------p    704
CONSENSUS-F   --------e--------S--------------K---------------------V----K------l--h    672
CONSENSUS-G   ?---------------?S--s--s---------------------k--------V---------------    662
CONSENSUS-O   v---?--kk--E--E--Si---l--TK-----K-A-I---A-?-?-V?MI?-NlVkNI----Q---L-iP    644
CONSENSUS-U   ?----?-??--?-----S-----?-??-----K----------------?-?-??-?-------------    593
CONSENSUS-CPZ ?---?-???-?E--?-?S-------T?-----K--?-?--?I?----?????-??R?-----?---?-??    355
                        <- tat cds                        *
CONSENSUS-A   tp?pr??pdRperIeeeGG.eQdrdRSirLVsGFLalAWDDLRsLCLfsYHrLRDfilI.a.ar.tVelL    714
CONSENSUS-B   l-a--g.-----g------.-r-----g---?-----i----------------Ll--?v?--?i----    733
CONSENSUS-C   --n--g----lg?------.-----------------------------------l---.-.--.a----    719
CONSENSUS-D   l-a--g.-----g------.--G--------n--s--i-----n----------l---.-.--.I----    720
CONSENSUS-E   ?HhQ-E.---------g--.--g----V-------------------------------.-.--.-----    768
CONSENSUS-F   i-S--E.-----g---G--.--Gk---V---n-----v-----N------rH-------.-.--.I-?..    734
CONSENSUS-G   ?HHQ-e.-----g---G--.-------------------------------------.-.--.-----    726
CONSENSUS-O   i?hq?E.a?T-G-TG-g--.-e--p-w?p?Pq---p-LYT---TII-W?--L-SNLaSg.I.........    696
CONSENSUS-U   ?----G.----G??-----.-?-??--?--??-----?-?------?---?----L?--.?.??.????-    636
CONSENSUS-CPZ ??-Q?-.?????E-?-?--.??--?-???-??-???-?-----N-GIW--QS-TSLACN.V.W-.QLKT-    396
                                              <- rev cds
                                                    ^^^
CONSENSUS-A   ghsslkglrlg.......weglkYL.wNLllyWgrELK?SAinLldtiAiavAgwtDRvIEigQrigRAi    775
CONSENSUS-B   -rr?....??-.......--a---w.----q--sq---n--vs--nat-----Eg------vv--a?---    787
CONSENSUS-C   -r---r--qr-.......--?-----.gs-vq---l---k---s----?-----Eg---i--?i---?---    777
CONSENSUS-D   -R.......R-.......--a----.----q--?q---n---S----------Eg---?--?v--a?--v    771
CONSENSUS-E   ---------R-.......-------.G-------Q---I---S--nat-------------vA-gaW---    830
CONSENSUS-F   ....?R---R-.......--A--l-.G--t----Q---N---s--N-T--v--Eg------AL--a----    791
CONSENSUS-G   -Rni-------.......-------.--------q---N-----?--------N-------vv--aC---    787
CONSENSUS-O   .qklIs?-g--LWIlGQk?IeaCR?caAvtQ--LQ--qn--T?----l-V---N---gi-lGi-----G-    761
CONSENSUS-U   ?????????R-.......--?----.G?-V?---?-------S-??AT--?--EG--??--??---?--?    675
CONSENSUS-CPZ --LI-HS---L.......R-R-CL-.GGIIQ---K---I---S---AT-----EG---I--AF-VTL-I-    458
CONSENSUS-A   lnIPrRIRQGlEraLl$                                                       791
CONSENSUS-B   -h--?------------                                                       802
CONSENSUS-C   ?---------f-a--q-                                                       792
CONSENSUS-D   -h--?------------                                                       786
CONSENSUS-E   -h---------------                                                       851
CONSENSUS-F   ----------------                                                        807
CONSENSUS-G   ----------------                                                        803
CONSENSUS-O   ?------------s---                                                       776
CONSENSUS-U   ?---------F-?--?-                                                       688
CONSENSUS-CPZ R---------------                                                        474
```

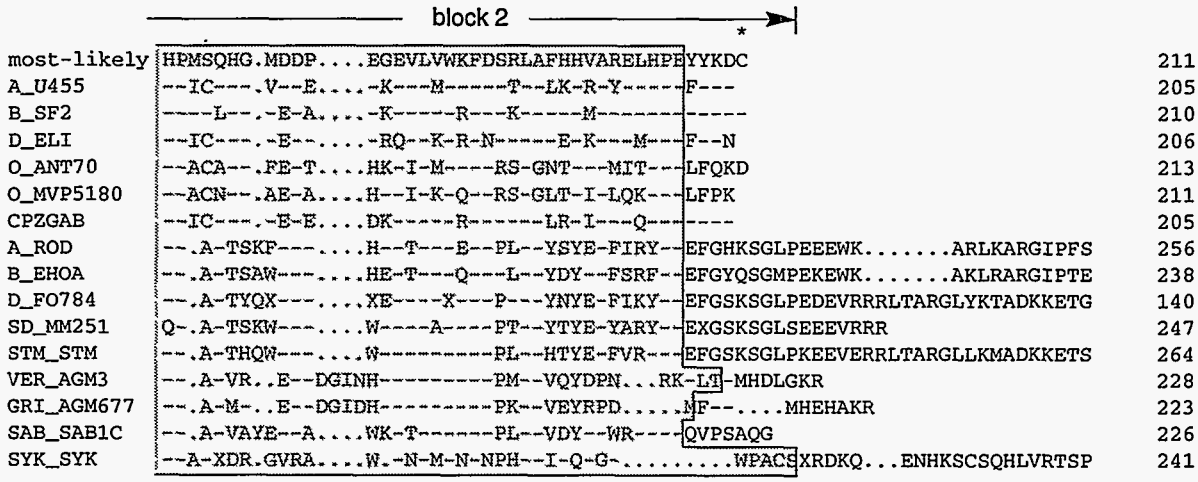
HMMER Sequences in the Nef Alignment

A_U455	HIVU455	M62320	Oram,J.D.	ARHR 6, 1073 (1990)
B_SF2	HIVSF2	K02007	Sanchez-Pescador,R.	Science 227, 484 (1985)
D_ELI	HIVELI	K03454	Alizon,M.	Cell 46, 63 (1986)
O_ANT70C	HIVANT70C	L20587	Vanden Haesevelde,M.	JVI 68, 1586 (1994)
O_MVP5180	HIVMVP5180	L20571	Gurtler,L.G.	JVI 68, 1581 (1994)
CPZGAB	SIVCPZGAB	X52154	Huet,T.	Nature 345, 356 (1990)
A_ROD	HIV2ROD	M15390	Clavel,F.	Nature 324, 691 (1986)
B_EHOA	HIV2EHOA	U27200	Rey-Cuille,M.A.	Virology 202, 471 (1994)
SD_MM251	SIVMM251	M19499	Franchini,G.	Nature 328, 539 (1994)
STM_STM	SIVSTM	M83293	Novembre,F.J.	Virology 186, 783 (1992)
VER_AGM3	SIVAGM3	M30931	Baier,M.	Virology 176, 216 (1990)
GRI_AGM677	SIVAGM677	M66437	Fomsgaard,A.	Virology 182, 397 (1991)
SAB_SAB1C	SIVSAB1C	U04005	Jin,M.J.	EMBO J. 13, 2935 (1994)
SYK_SYK	SIVSYK	L06042	Hirsch,V.M.	JVI 67, 1517 (1993)

The following alignment and most-likely sequence were generated using the HMMER program as described in Part III. For simplicity, only representative types and subtypes are shown. The annotation is based on HIV1s, therefore the user should be cautious about its applicability to other PIVs. Cysteines are indicated by '*'. Two conserved gapless blocks were found using the BLOCKMAKER/Motif program; these are shown with shading. The Motif "Cobbler" sequence follows this alignment. For further information about Nef, see the section by S. Wain-Hobson in Part II of this compendium.

```
most-likely MGGKWSK....RSK.SGWPAVRERMRRAEPET.....YGPAADGVG.AVSR.DLEKHGA.........ITSS..     51
A_U455      --------....K-R.VE--E--K---ETPA..........--K---.---Q.--D-Y--.........V---..     46
B_SF2       --------....--M.G--S-I---------.R.....AE-------.----.-------.........----..     50
D_ELI       --------....S-I.V----I---I--TN-.....------.----.-------.........----..     46
O_ANT70     --NALR-.....G-FE--A--------TRTFPE....SE-C-P---.QI--.E-AAR-G.........-P--..     52
O_MVP5180   --NA---.....--FA--SE--D----SSSDP.....QQ-C-P---.-----.E-ATR-G.........-S--..     51
CPZGAB      --T----....S-L.V---E--R-I-E-PT..........--E---.E--K.---R---.........---R..     46
A_ROD       --ASG--...KH-RPPRG..LQ--LL--RAGA.....C-GYWNES-GEY--.FQ-GSDR.........EQK-PS     54
B_EHOA      --SAG--...KQ--QQPG..L---LL--RRGPR....GESSGERQE.RSLQ.YPGGSDK.........GLN-PS     54
SD_MM251    ---AI-M...R---PA-D..L-QKLL--RG--.....--RLLGE-EDGS-Q.S-GGL-K.........GL--RS     54
STM_STM     --ASG--...KQR-QH-E..L---LL--RG--.....--KLLE-L-EGSGP.SQGASDK.........GLN-HS     54
VER_AGM3    --LGN--PQHKKQL.-L-H-LH....KTRATR.....--LL--PLIGQS-T.LQ-ECDK.........ALKESL     54
GRI_AGM677  --SSN--.....-QQ.Q-LLKLW.-GL-GK-GAD....WVLLS-PLIGQS-T.VQ-EC-K.........ALKK..     52
SAB_SAB1C   ----S--...QQQRH-L-..LWSKL-Q-PVIQ.....-DML--PLLGQS-H.IQ-ECAKSLRDGLIRQGD--..     61
SYK_SYK     --STS--.........-QQ..L-SEGKY-IGWRLFGKQ-T-LP-ELS.RPLQPCRGGFDK.........AWR-TL     54

                                      *
most-likely ............NTA.ANNADCAWL..EAQ..........E..EEEVGFPVRPQVPLRPMTYKAAVDLSHFLKE     98
A_U455      ............--S.ST--S----..---.........-..-GD-------------------F---F----     93
B_SF2       ............---.-T-------..---.........-..----------------------L-I------     97
D_ELI       ............---.ST-------..---.........-.ESD------------------E-L--------     94
O_ANT70     ............H-P.Q---AL-F-..-SH..........Q..--------A-----------G-F---F----     99
O_MVP5180   ............H-P.Q---AL-F-..DSH..........K..D-D---------------F---F---F----     98
CPZGAB      ............--P.ET-QTL---..-EM..........D..N-----------T--------F--------     93
A_ROD       CEGRQYQQGDFM--PWKDP-AEREK..NLYRQQNMDDVDSD.DDDQ-RVS-T-K-------HRL-I-M--LI-T    125
B_EHOA      CDGQ........K-L......G-EGG.GK-....DSDEDD-..DN---VR---G-------F-L---M------    107
D_FO784                                                                          X-       2
SD_MM251    CEGQKYNQGQYM--PWR-P-EEKEK..L-YRKQNMDDIDE-.DDDL--VS---K----A----L-I-M---I--    125
STM_STM     CEPQRYNEGQFM--PWK-P-AESAK..LEYRQQNMDDVDE-.DDNL--VA-H-R----E----L-I-----I-S    125
VER_AGM3    IRK.....RNGKM-P.....EGRK-Q.-GD....KWDEWSD.E-D--------R----Q----L---F------    112
GRI_AGM677  ......SWGKGKM-P.....-GRR-Q.-GD....TFDEWDD.D--------Q-R----Q----L---F-----S    109
SAB_SAB1C   ............R-E.....EGVKM..KH-GRQP..SWYD-.D----------CL---A----L-I-FG-----    113
SYK_SYK     TEP......IDPHGP.....-RD-GHSGG-KFSPGDIVQD-GDTGL-----C--T---TL---L-I-----I-N    117

                                  *                                    *
                    block 1 ◄──────────────────────────►                  ◄── block 2
most-likely KGGLEGLIYSQRRQDILDLWVYHTQGYFPDWQNYTPGPGIRYPLTFGWCFKLVPVEPEKVEEA.NEG.ENNCLL    170
A_U455      ----D---H--K--E----------F--------------------Y-----D-AE-----.TG-.---S--    165
B_SF2       --------W-----E-----I-------------------------------.---.---S--    169
D_ELI       --------W-KK--E-------N---I----------------------YE----D-QE---D.T--.-T-S--    166
O_ANT70     ---------HK-AE-------N---F----------T-F------L------SE-EA-RLG-TC.-RAN--    172
O_MVP5180   ----D---HK-AE-----I-----F------C-------P-F-------L-------SA-EA-RLG-TN.-DAS--    171
CPZGAB      -------V--R---E----------F--------T---T-F--C---------LTE-Q--Q-.---.D-----    165
A_ROD       R-----MF--E--HK--NIYLEKEE-IIA-------H---V---MF---LW-----DVP..Q-G.EDT.-TH--V    195
B_EHOA      --E---IF--E--HK---TYLENEE-IVSG-----H---V---KF---LW----INMI..A-P.EDE.-TH--V    177
D_FO784     ------IY-NH--HR---IYLQNEE-II-------S---E-L-MMY--LW-----DVP..D--.QGD.-RH--V     72
SD_MM251    ------IY--A--HR---MYLEKEE-II----D--S-------K----LW-----NVS..D--.Q-D.-RHY-M    195
STM_STM     ------IY--E--HR---MYLEKEE-IV-------A-------KQ---LW-----DMS..N--.Q-DDGTHY-V    196
VER_AGM3    ----D-IY--D--NQ--N-YALNEW-IID--NAWSE-------RC--F-------DLH..----.ETC.-RH--V    182
GRI_AGM677  ----D-IY--E--EK--N-YALNEW-IID---A-S--------RV--F-------DLH..----.RNC.-RH--M    179
SAB_SAB1C   ------IY--E--KK----YALNEW-IVDG-----D---T---KC----------DLS..----.KNS.--H---    183
SYK_SYK     ----Q-MN-CEK-DE--H-YLQNEH-II.-RI---S---T-----I---LWE---N-I-..GCL.EYE.-HTL--    186
```

HIV1 NEF



COBBLER sequence from MOTIF

>NEF STM_STM, with embedded consensus blocks

mgasgskkqrkqhgelrerllrargetygkllleglgegsgpsqgasdkglnshscepqry
negqfmntpwknpaaesakleyrqqnmddvdeeddnlvgvavhprvplremtyklaidls
hfikskgglegiyYSERRHEILDLWVYNEEGIIPDWQNYTPGPGVRYPLTFGWCWKLVPV
Dmsneaqeddgthyllhpacthgxddphgevlvwkfdprlayeyvarxkhpeefgsksgl
pkeeverrltargllkmadkkets

```
CONSENSUS-A MGGKWSKsSiVgWPeVrkRmRqT.............?PtAAkGVGAvSQD.....LDkhGAiTSSNt??     48
CONSENSUS-B -------?-??---?--e---ra??????????????-Ep--d------r-.....-e----------aa     46
CONSENSUS-C --------------A--E-I-R-.............QP--E----A---.....---Y--L-----PT     50
CONSENSUS-D --------------AI-E-I-r-?????.........dP--D------R-.....-E----------as     50
CONSENSUS-E --------------Q--E-ik--.............-p-te--------.....------v----m..     48
CONSENSUS-F --?-----------AI-E-L-?-.............-P--E--------.....-ERR----?---R-     46
CONSENSUS-G --------------A--E-I-R-PPAAERK........EA--E--------.....-AR---V-----AA     57
CONSENSUS-O --NA??-?KF?--??--?---R?.........???P?-?PC-P---??-RE.....-A?R-G-?--H-PQ     38
CONSENSUS-U --?----????---??-E-I-?-???.............P???----?---?????-?-?--??--??A-     31

               *       SH3-binding |  |  |  | SH3-binding
CONSENSUS-A tnpsCaWLE?Aqe?.d..e?.VGFPVRPQVPLRPMTYKgAvDLShFLKEKGGLDGLIyS?kRQEILDLWV    110
CONSENSUS-B --ad-----.----.e??-e?-----------------a-?------------e---?-q---d------    108
CONSENSUS-C S-AD----Q.---..EE.GE.-------------------F---F-                            91
CONSENSUS-D --ad-----.---..ES.-E.-----------------e--------------E---W-K----------    115
CONSENSUS-E n-aD-v--r.---..e??-g.-------------------f---F--------------Kr---------    112
CONSENSUS-F N--DL----.---..E..?E.-------------?-----?------------E-----K?---------    106
CONSENSUS-G N--D-----.---..EE.SE.---------------F-S-F---F-                            98
CONSENSUS-O N-AAL-F-?.SH?..?..--.-----?---------?-?-F---F--------?-----H--A------?     93
CONSENSUS-U N-??-???-.??-..E?.-E.---------------?---F---?-----------??------------     83

                                          SH3-binding
                                        *     |  |
CONSENSUS-A YnTQGfFPDWQNYTPGPGtRf.PLTFGWCfKLVPvDPaEVE.eat?GEnNSLLHPICQHGmdDe?revLm    176
CONSENSUS-B -h---y------------?-y?-------------e-ek--.--ne---------msl-----pE----?    174
CONSENSUS-D -----I------------I-Y.--------e------q---.---E--t-c----?-----E-pE-q--k    182
CONSENSUS-E ------------------i-y.--C------------r---.-dnk----C----ms---ie--e-----    180
CONSENSUS-F -H                                                                        108
CONSENSUS-O -?---------?------?--.------L------S?E-A-RLGNT?-?A?----A-?--?E-?H?-I-?    150
CONSENSUS-U -H---?----?-------?-?.--?---------??-?---.--N-----C----?S----?-?E----?    138

CONSENSUS-A WkFDSrLAlkHrA?ElHPEfY.KDC$                                               199
CONSENSUS-B -r------fh-m-r-----y-.---?$HRCVYKGLSAGAFPGRRGWAGLGSGEPSDAA               229
CONSENSUS-D -R-N----fE-K-R-m-----.---                                                 206
CONSENSUS-E -----a--R--i-R-----y-.---                                                 204
CONSENSUS-O -?--RS-G?T-?--??---LF?-?                                                  166
CONSENSUS-U -----S--??-?-R-?---?-.---                                                 157
```

III

Analyses

PART III Analyses

HIV-1 Tat: Structure and Function

Kuan-Teh Jeang

Molecular Virology Section, LMM, NIAID, National Institutes of Health
Bethesda, MD 20892-0460, USA

Introduction

Most viruses encode functions for regulating genome transcription. Examples include the SV40 T-antigen, the adenovirus E1a protein, and the herpes virus immediate-early proteins. For the human immunodeficiency viruses (HIV), Tat functions similarly, though not identically, to those other activators. Over the past decade, we have learned much about Tat, both in structure and in function. The intent of this brief accompanying overview is to "add flavor" to raw data, it is written for the purpose of apprising, in a short format, the readers on some of the current thoughts about Tat. For more in-depth discussions, extensive subject reviews can be found elsewhere (e.g. Jones and Peterlin, 1994; Chang et al., 1995; Gatignol et al., 1996). This overview was revised from an earlier version appearing in the 1994 compendium.

Tat function

Tat is a small nuclear protein of 86 to 101 amino acids (depending on the viral strain) which is encoded from two separate exons (see Section I). Analyses of "full-length" Tat have been performed commonly using the 86 amino acid version. However, it should be noted that while a few laboratory strains (e.g. HXB2 and NL4–3) have the truncated Tat (86 aa) most HIV-1s have the 101 aa protein (see compendium part II).

A. LTR Transcription. Despite intensive efforts, the mechanism of Tat action remains incompletely understood. It is accepted that Tat is required for optimal HIV viability (Fisher et al., 1986; Dayton et al., 1986). The role of Tat in critically directing transcription (Peterlin et al., 1986; Rice and Mathews, 1988; Laspia and Mathews, 1989) from the HIV LTR is one necessary function suggested for this protein. However, increasingly there is evidence that Tat has other significant effects on the virus and on the host cell (Huang et al., 1994; Chang et al., 1995; Neuveut and Jeang, 1996; Goldstein, 1996). In the setting of viral pathogenesis, both aspects need to be considered.

In activating transcription from the LTR, Tat differs from other prototypic viral trans-activators in requiring a bipartite responsive element consisting of DNA and RNA. To our knowledge, Tat is the first characterized eukaryotic transcription factor that binds to a nascent leader RNA, TAR (Berkhout et al., 1989; Dingwall et al., 1989; Cordingley et al., 1990; Roy et al., 1990; Calnan et al., 1991); and then influences events at the TATAA-enhancer-promoter (Berkhout et al., 1990; Selby and Peterlin, 1990; Southgate et al., 1990; Jeang et al., 1993b; Gatignol et al., 1996; see fig. 1A). TAR RNA has an extensive secondary structure including a stem, a bulge, and a loop (Muesing et al.,1987; Berkhout and Jeang, 1989; Roy et al., 1990c; Wang and Rana, 1996; see fig. 1B). Early studies indicated that the UCU sequence of the bulge is critical for binding by Tat (Dingwall et al. 1989,; Roy et al. 1990a, Calnan et al., 1991; Cordingley et al., 1990). By contrast, the structure of the stem, but not its specific sequence, was proposed to be important for function. The loop of TAR RNA serves as binding site(s) for cellular factor(s) that cooperate with Tat in the activation of the LTR (Sheline et al., 1991; Wu et al., 1991). A more extensive discussion of the role of cellular factors that bind TAR RNA (Gatignol et al., 1989; Gaynor et al., 1989; Gatignol et al., 1991; Gatignol et al., 1996) is presented in Section IV.

Both for Tat binding and for transcription, there are sequence specific requirements for the immediate stem nucleotide pairs that flank the bulge (Weeks and Crothers, 1991; Berkhout and Jeang, 1991; Churcher et al., 1993; Wang et al., 1996). In a recent series of studies, Rana and colleagues have provided physical insights on how Tat interacts with TAR RNA. Using photo-activated cross-linkers, they were able to elucidate the orientation of Tat as it intercalates into TAR RNA. They found that

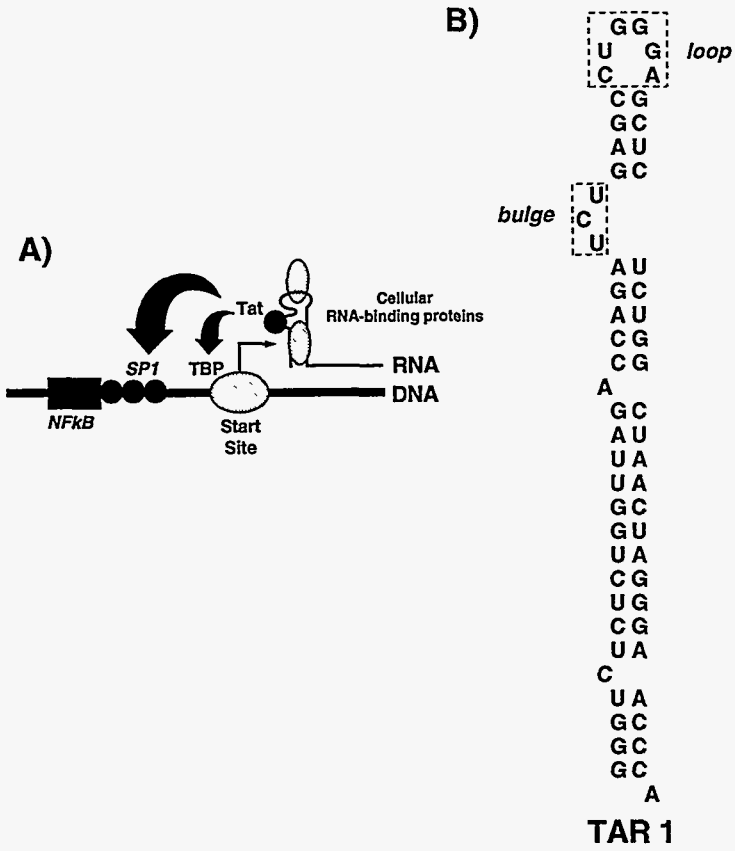


Fig. 1. Interaction of Tat with DNA and RNA targets in the HIV-1 LTR. A) A schematic representation of the functional interactions between Tat, TAR-RNA-binding proteins and promoter elements. Biochemical evidence exists that Tat contacts directly SP1 (Jeang et al., 1993b) and TATAA-binding protein (TBP; Kashanchi et al., 1994). B) Secondary structure of TAR RNA. The crucial trinucleotide bulge and hexanucleotide loop elements are boxed.

amino acid 41 of Tat lies in close proximity to U42 in the lower TAR stem, that amino acid 47 of Tat is proximal to the G26 nucleotide positioned immediately above the TAR bulge, and that amino acid 57 of Tat is close to U31 in the TAR loop (Wang and Rana, 1995; Liu et al., 1996; Wang et al., 1996). At the same time, the basic domain of Tat was shown to interact with RNA residues U23, U38, and U40 and to distort/widen the major groove of TAR RNA (Wang and Rana, 1996; Metzger et al., 1996).

While it is clear that Tat binds TAR RNA and interacts with enhancer-promoter-binding factors (Berkhout and Jeang, 1992; Jeang et al., 1993a; Southgate and Green, 1995), the direct mechanism through which these physical events influence transcription is not wholly evident. A number of models that attempt to explain the transcription function of Tat have been proposed. These include i) anti-terminating (Kao et al., 1987) stalled RNA polymerase II (RNAP II), ii) increasing processivity/elongation of transcribing RNAPII complex (Laspia et al., 1989; Marciniak et al., 1990; Hermann and Rice, 1995; Zhou and Sharp, 1996), and iii) facilitating formation of RNAPII complexes at the promoter (Laspia et al., 1989, Jeang and Berkhout, 1992; Jeang et al., 1993a; reviewed in Cullen, 1993). Currently, there are findings consistent with each of these models. Possibly, Tat, like basal transcription factor TFIIF (Buratowski, 1994), acts simultaneously as an initiation and elongation factor. Alternatively, because transcription from a given promoter is a dynamic multi-cycled process, efficient

disassembly (i.e. processive promoter-clearance) of a previous round is necessary to permit initiation of the next round (reviewed in Zawel and Reinberg, 1995). Thus, initiation and elongation/clearance events at the promoter may be mechanistically linked and, in some cases, inseparable processes. Recent findings indicate that Tat plays a mechanistic role not at the stage of recruiting TBP to the promoter (Kashanchi et al., 1994), but at the step of clearing RNA polymerase II from the promoter (Chun and Jeang, 1996; Xiao and Jeang, unpublished observations).

Lastly, one must exercise care in not assuming that different look-alike experimental systems would share the same the rate-limiting step (Jeang et al., 1993a). Two examples suggest caution. First, results from Pomerantz and co-workers have recently shown that short LTR-transcripts are not found in cells that are fully functional for Tat-transactivation (Niikura et al., 1996). Second, some of the analogies drawn between EIAV Tat and HIV-1 Tat bear re-visiting. One recent report has raised the consideration that EIAV Tat is a DNA-binding and not an RNA-binding protein (Rosch and Willbold; 1996).

B. Translation. There are suggestions that Tat also functions in regulating translation (Rosen et al., 1986; Cullen, 1987). TAR RNA can inhibit translation (Parkin et al., 1988; SenGupta and Silverman, 1989) of HIV-1 mRNAs, most likely through activation of double-stranded RNA-dependent protein kinase (PKR) and 2–5A synthetase (SenGupta and Silverman, 1989; Edery et al., 1989). Addition of Tat was found to reverse this translational inhibition (SenGupta et al., 1990; Braddock et al., 1990). It is possible that one part of Tat's translational effect stems from its ability to form physically a functional complex with PKR (McMillan et al., 1995).

C. Effects on cellular function. There is increasing evidence that Tat has pleiotropic effects on cellular genes and host cell metabolism (reviewed in Chang et al., 1995). This would not be surprising, since it clearly serves the advantage of the virus to be able to modify optimally the cellular environment for replication. In this regard, Tat has been reported to function as a secreted growth factor in stimulating the growth of Kaposi-like cells (Ensoli et al., 1990; Ensoli et al., 1993; Barillari et al., 1993) and in promoting angiogenesis (Albini et al., 1996). Tat is further described to affect the organization of neurons and astrocytes (Kolson et al., 1993); it is neurotoxic at low concentrations (Sabatier et al., 1991).

Expression of many cytokines is modulated by Tat (Rautonen and Rautonen, 1992). These include TNFa,b (Buonogaro et al., 1992; 1994), TGFa,b (Lotz et al., 1994; Nabell et al., 1994), IL-2 (Purvis et al., 1992; Westerndorf et al., 1994), and IL-6 (Scala et al., 1994) among others. Tat has also been reported to activate cellular signal transduction pathways that involve phosphatidylinositol-3-kinase (Milani et al., 1996) and NF-kB (Biswa et al., 1995). It has been further suggested that Tat can affect programmed cell death by protecting lymphocytes against apoptosis (Gibellini et al., 1995). However, the validity of this last point point has been contested by conflicting findings that Tat promotes apoptotic death of lymphocytes (Li et al., 1995).

Domains in Tat

Tat is synthesized from an mRNA joined from two coding exons. The first exon encodes amino acids 1–72 and (in most strains of HIV-1) the second exon encodes amino acids 73–101 (see fig. 2). In reporter plasmid co-transfection assays, the first 72 amino acids of Tat fully trans-activates transcription from the LTR. In fact, a truncated 58 amino acid form of Tat is virtually wild type in this type of co-transfection assay (Seigel et al., 1986; Garcia et al., 1988; Kuppusway et al., 1989). Whether data from plasmid co-transfections reflect accurately the normal physiological function of Tat during replication of HIV-1 in lymphocytes has been a poorly studied issue. Recent investigations suggest that reporter plasmid co-transfection results cannot be interpreted literally for their relevance towards viral replication (Neuveut and Jeang, 1996).

A. First coding exon. The combined results from many laboratories have permitted an arbitrary demarcation of "domains" in Tat (Kuppuswamy et al., 1989). For instance, the N-terminus of Tat

(domain 1; .fig. 2) has 13 amino acids with amphipathic characteristics. Mutations that alter the acidic composition of this region were felt originally to affect trans-activation (Rappaport et al., 1989); however, results from a later study conflicted with this interpretation (Tiley et al., 1990).

Amino acids 22 to 37 (domain 2, fig. 2) contain seven cysteines and are highly conserved between different isolates of HIV-1s. Individual mutation in six of the seven cysteines abolish Tat function (see Table I). Although originally proposed as a metal-chelating dimerization domain (Frankel et al., 1988), this region was recently shown to be used for intra-molecular disulfide bond formation in monomeric molecules of Tat protein found inside cells (Koken et al., 1994). Currently, it is believed that Tat is active functionally as a monomer rather than a dimer (Rice and Chan, 1991; Koken et al., 1994).

Domain 3 (amino acids 40 to 48) contains a RKGLGI motif that is conserved between HIV-1, HIV-2 and SIV Tat. This region, in conjunction with the amino terminus and the cysteine domain, has been suggested to circumscribe the minimal activation domain of HIV-1 Tat (Carroll et al., 1991; Derse et al., 1991). Domain 4 (amino acids 49–72) contains a basic RKKRRQRRR motif. These amino acids confer TAR RNA-binding properties to Tat (Dingwall et al., 1989; Roy et al., 1990; Weeks et al., 1990; Chang and Jeang, 1992) and are important for nuclear localization of the protein (Ruben et al., 1989; Hauber et al., 1989). However, recent studies suggest that this short basic stretch is insufficient in determining the entire specificity of Tat-TAR binding since amino acids outside of the basic domain also contribute to this interaction (Churcher et al., 1993; Luo et al., 1993).

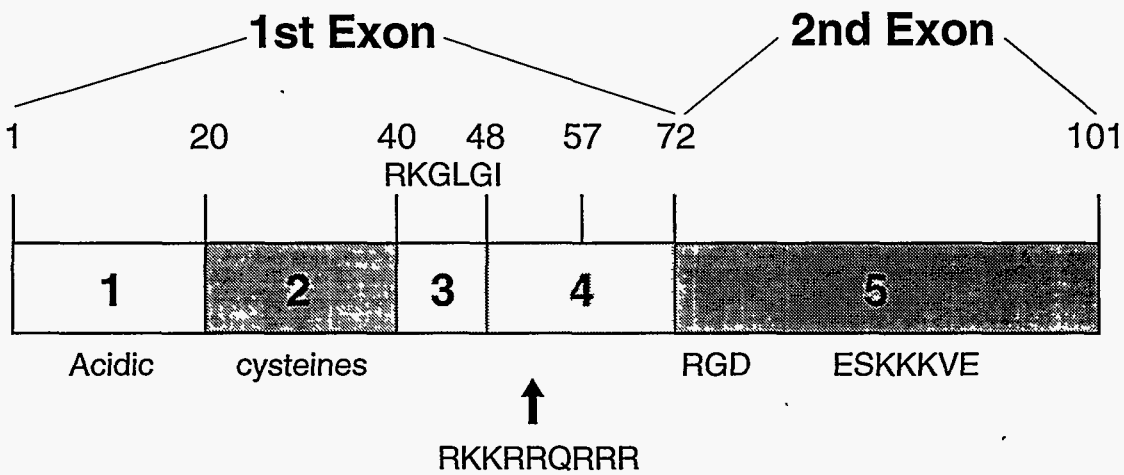


Fig. 2. Domain classifications of Tat protein. The demarcation of domains is somewhat arbitrary. The first exon includes amino acids 1–72, while the second exon includes 73–101. Motifs and characteristics of each "domain" are indicated above or below each region.

Table I summarizes 77 point mutations in Tat collated from the work of nine laboratories (Garcia et al., 1988; Sadaie et al., 1989; Kuppuswamy et al., 1989; Ruben et al., 1989; Hauber et al., 1989; Meyerhans et al., 1989; Rice and Carlotti, 1990a; Rice and Carlotti, 1990b; Siderovski et al., 1992; Neuveut and Jeang, 1996). As alluded to above, most studies were performed based on measurements from reporter-plasmids. Because the first 58 amino acids of Tat recapitulate well the trans-activation function in this type of assay, it is not surprising that the great majority of engineered mutations is concentrated within amino acids 1 to 58. In many cases, individual amino acids have been changed to more than one counterpart, heightening the validity of the resulting phenotype. In rare instances, mutational analysis of the second coding exon of Tat has been studied. Two recent point mutants of the second exon of Tat (P81fs and K90R) showed unexpectedly that changes in this exon perturb not measurements of plasmid trans-activation but do affect virus replication (Neuveut and Jeang, 1996).

Table I. Point mutations in Tat

Amino Acid Changes From	To	Activities	Consensus Amino Acid
Q2	A	++	E
P3	A	++	P
P3	Q	++	P
V4	A	++	V
D5	A	+	D
P6	A	++	P
P6	S	++	P
Δ3–6		+	
P6P10	LL	++	
R7	A	++	R
L8	A	++	L
E9	A	++	E
P10	A	++	P
P10P13	LL	++	
W11	A	++	W
K12	A	++	K
K12	N	++	K
P18	A	++	P
K19	R	++	K
A21	D	++	A
A21T23	VA	+	
T23	A	++	T
C22	S	–	C
C22	G	–	C
N23	T	++	T
N24	A	++	N
N24	K	++	N
C25	R	–	C
C25	G	–	C
Y26	A	+	Y
Y26	F	++	Y
C27	S	–	C
C27	G	–	C
K28K29	AA	+	KK
K28K29	EA	–	KK
C30	G	–	C
C31	S	++	C
C31	E	–	C
C31	G	++	C
F32	A	+	F
H33	A	–	H
C34	G	–	
C34	S	–	C
G35	A	+	Q
C37	G	–	C
C37	S	–	C

Table I. (cont) Point mutations in Tat

Amino Acid Changes From	To	Activities	Consensus Amino Acid
F38	A	−	F
F38	L	++	F
K40	D	−	T
K40	T	++	T
K41	A	−	K
*K41	T	−	K
*K41	T	++	K
L43	F	+	L
G44	S	++	G
S46	A	++	S
S46	P	−	S
Y47	H	++	Y
Y47	A	++	Y
G48	S	++	G
G48R49SG[1]		++	
R49	T	++	R
K50	stop	−	K
K50K51	Y50Y51	+	KK
K50K51	S50G51	+/±	KK
K50	E	++	K
K50	T	++	K
R52	E	++	R
R53	I	++	R
Q54	N	++	Q
R55	G	+	R
R55R56	L55T56	+	
R56	E	++	R
R57Q63	SE	++	RQ
L69	I	++	L
P81	fs[2]	−[3]	P
K90	R	−[3]	K

Column 1: First letter indicates original amino acid. Number indicates position of amino acid in Tat.

Column 2: Letter(s) indicate the resulting amino acid.

Column 3: ++ > 50% wild type activity; + > 10% wild type activity; +/− or − indicate < 10% wild type activity.

Column 4: Because not all isolates of HIV-1 have the same amino acids for Tat, a consensus sequence is listed also.

Notes: * Different results reported for the same mutation from Kuppuswamy et al., 1989; and Meyerhans et al., 1989. [1]Amino acids beyond position 59 completely changed. [2]fs = frame shift of amino acids beyond position 81. [3]Measurement is based not upon trans-activation of a reporter plasmid but on delayed replication of an HIV-1 molecular clone in T-cell lines (Neuveut and Jeang, 1996).

Examination of the mutants (Table 1) reveals that the region spanning amino acids 1–21 is remarkably tolerant of changes. In contrast, changes in amino acids 22 through 40 were generally deleterious for trans-activation. Finally, although the basic domain (amino acids 49–57) as a unit is necessary for Tat function, individual amino acid changes do not significantly affect activity.

B. Second coding exon. The second coding exon of Tat has been less studied. In routine transfection of reporter plasmids, absence of the second exon does not alter greatly measurements of Tat activity in this type of assay. However, findings from HIV-2 and SIV Tat are quite clear in demonstrating that this exon contributes towards optimal trans-activation (Viglianti and Mullin, 1988; Tong-Starksen et al., 1993). There are emerging findings that the second exon of HIV-1 Tat, in other assays, is important for trans-activation (Jeang et al., 1993b), trans-repression (Howcroft et al., 1993), and virus replication (Neuveut and Jeang, 1996).

Two short motifs in the second exon of HIV-1 Tat could have functional importance (see fig. 2). The first is an RGD sequence that is used as a cell adhesion signal for binding to cellular integrins (Brake et al., 1990). This RGD motif , however, is not found in HIV-2 or SIV Tat proteins. Recently, it has been suggested that the basic domain of Tat (rather than the RGD sequence) is the important protein portion for cellular uptake (Albini et al., 1996). The second exon also has an ESKKKVE motif which is conserved in most HIV-1 Tat proteins and is partially preserved in HIV-2 and SIV Tats. The functional significance of this motif has not been examined in detail.

C. Immunological epitopes in Tat. Evidence suggests that Tat has a role in viral infectivity and can contribute to pathogenesis (Huang et al., 1994; Neuveut and Jeang, 1996). Indeed, in tissue culture experiments, findings suggest that the addition of antibody against Tat to the culture supernatant can halt the spread of HIV-1 infection in T-lymphocytes (Steinaa et al., 1994; Re et al., 1995). Complementary observations in patients indicate that immune response to Tat correlates with better prognosis for disease progression (Reiss et al., 1991). These results have led to the suggestion that Tat should be considered as a candidate prophylactic vaccine (Goldstein, 1996) for HIV-1.

Although the immunological epitopes within Tat have not been studied in detail, some information is available. Figure 3 summarizes findings from 4 studies. Two regions of Tat (amino acids 1–9 and 70–83) have been found to be potent in eliciting humoral responses. Three other regions have been described to elicit CTLs. Interestingly, it has been observed that uninfected individuals have natural IgM antibodies directed against two portions of Tat, suggesting that this might be one basis for natural immunity against the virus.

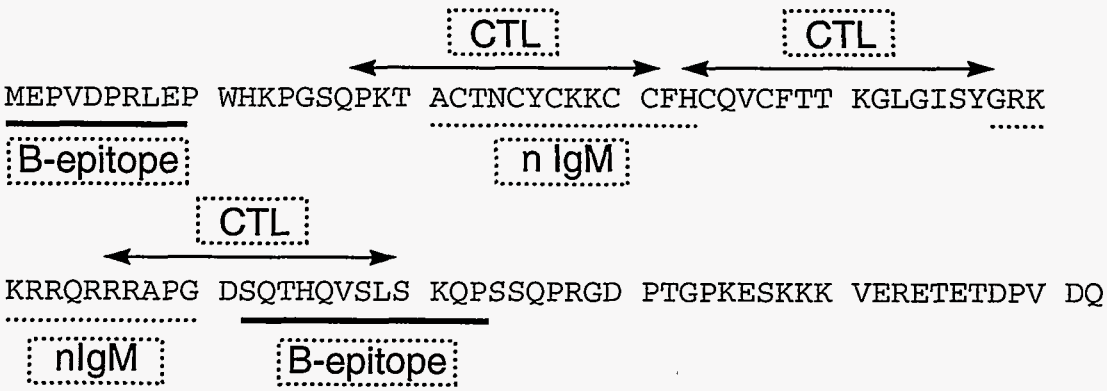
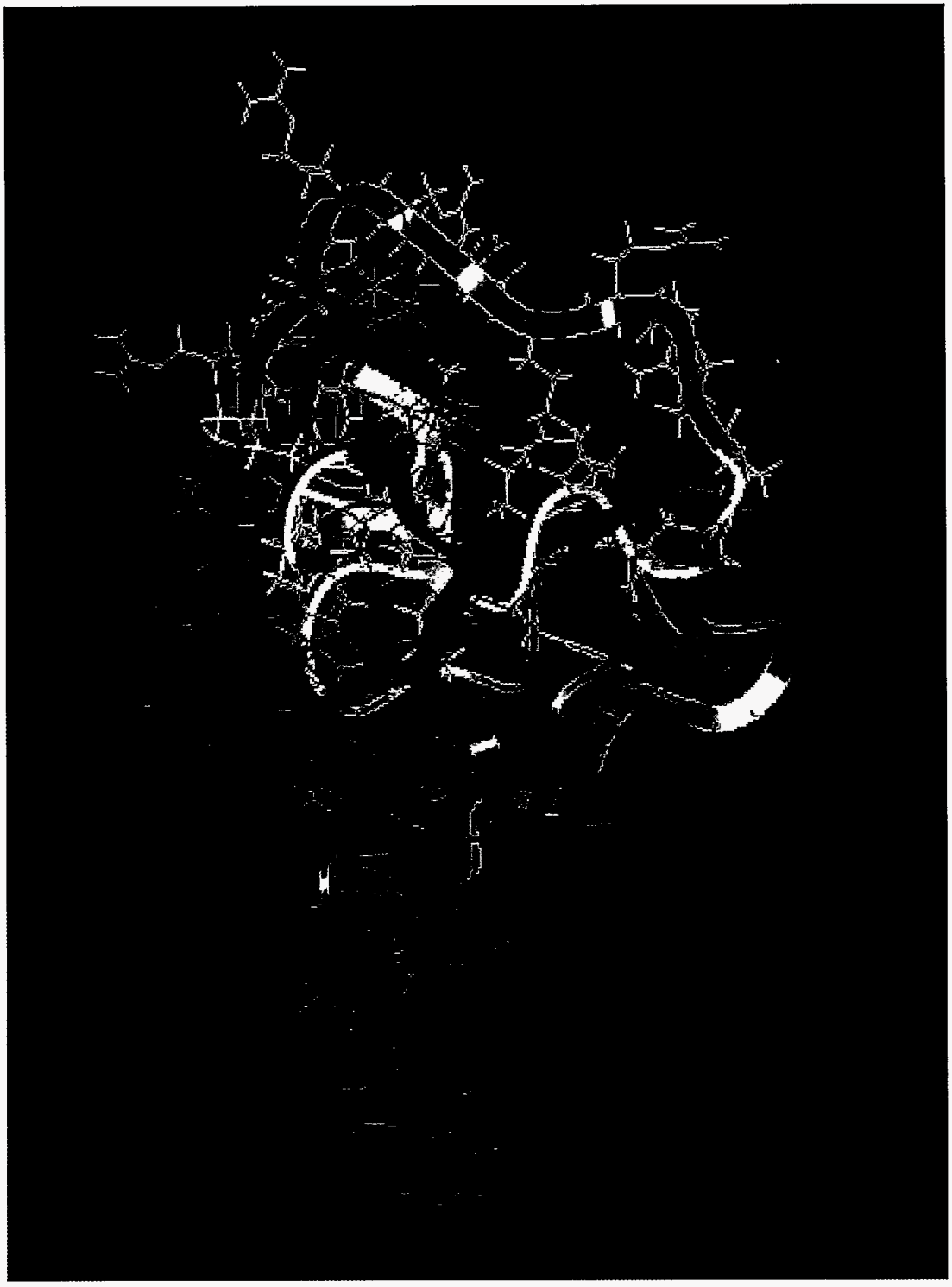


Fig. 3. Immunological epitopes in the HIV-1 Tat protein. B-cell immunodominant epitopes were determined by Krone et al. (1988) and McPhee et al. (1988); CTL-epitopes were mapped by Blazevic et al. (1993); and naturally occurring IgM epitopes in Tat were as reported by Rodman et al. (1993).

Concluding Remarks

For a relatively small protein, much effort has been expended on studying Tat. A recent search of Medline (conducted Nov. 1996) using the word Tat yielded more than 1400 papers in the last three year period. Obviously, one can't cover all these new findings in this brief summary on this intriguing protein. It is hoped that this synopsis highlights some of the currently important issues on Tat, albeit considered in a rather subjective manner. I hope to update periodically the contents in this text in coming years and in coming editions of the database.

I conclude this writeup with a figure (figure 4, provided by Dr. E. P. Loret) showing recently derived computer models of Tat structure (Gregoire and Loret, 1996). As yet, a crystal structure for Tat is not available. However, analyses of multiple sequences do yield informative findings. For instance, these analyses reveal that the regions circumscribed by amino acids 38–47 and 59–72 have the widest degree of three-dimensional structural variability between different HIV-1 Tat proteins. The structure of other portions of Tat seems to be well-conserved. At the level of our current understanding of Tat function, it is difficult to establish a clear one-for-one correlation between the physical model (figure 4) and the functional phenotypes revealed by point mutants (Table 1). In coming years, I hope that this will change.



Tat BRU structure obtained from molecular modeling (Gregoire and Loret, 1996) using the atomic coordinates of the Tat Z2 2D-NMR structure (Bayer et al., *J. Mol. Biol.* 247:529–535, 1995). Region I is colored in red, region II in orange, Region III in yellow, region IV in green, region V in light blue and region VI in blue. Molecular modeling was made with Insight II, Discover and Homology from MSI Technologies, Inc. (San Diego, CA) running on a R4600 Silicon Graphics Workstation. The CVFF force field was used to minimize the structure. Hydrogens were generated at pH 7. Steepest descent and conjugate gradient were the algorithms used for the minimization. Dynamic was performed at 300 K and 110 different structures were analyzed in the trajectory.

Backbone superimposition of Tat Bru (green) and Tat Oy (pink). The C-terminal extremity of Tat Oy (yellow) goes through the loop made by the cysteine rich region (region II) and ends up in a groove made by a part of region I and region III.

References

[1] Albini, A., R. Benelli, M. Presta, M. Rusnati, M. Ziche, A. Rubartelli, G. Paglialunga, F. Bussolino, and D. Noonan. 1996. HIV-tat protein is a heparin-binding angiogenic growth factor. *Oncogene.* **12**:289–297.

[2] Barillari, G., R. Gendelman, R. C. Gallo, and B. Ensoli. 1993. The tat protein of human immunodeficiency virus type 1, a growth factor for AIDS Kaposi sarcoma and cytokine-activated vascular cells, induces adhesion of the same cell types by using integrin receptors recognizing the RGD amino acid structure. *Proc. Natl. Acad. Sci. USA* **90**:7941–7945.

[3] Berkhout, B., A. Gatignol, A. B. Rabson, and K.-T. Jeang. 1990. TAR-independent activation of the HIV-1 LTR: evidence that Tat requires specific regions of the promoter. *Cell* **62**:7257–7267.

[4] Berkhout, B., and K.-T. Jeang. 1989. trans Activation of human immunodeficiency virus type 1 is sequence specific for both the single-stranded bulge and loop of the trans-acting-responsive hairpin: a quantitative analysis. *J. Virol.* **63**:5501–5504.

[5] Berkhout, B., and K.-T. Jeang. 1991. A detailed mutational analysis of TAR RNA; critical spacing between the bulge and loop recognition domains. *Nucleic Acids Res.* **19**:6169–6176.

[6] Berkhout, B., and K.-T. Jeang. 1992. Functional roles for the TATA promoter and enhancers in basal and Tat-induced expression of the Human Immunodeficiency virus type 1 long terminal repeat. *J. Virol.* **66**:139–149.

[7] Berkhout, B., R. H. Silverman, and K.-T. Jeang. 1989. Tat trans-activates the human immunodeficiency virus through a nascent RNA target. *Cell* **59**:273–282.

[8] Biswas, D. K., T. R. Salas, F. Wang, C. M. Ahlers, B. J. Dezube, and A. B. Pardee. 1995. A Tat-induced auto-up-regulatory loop for superactivation of the human immunodeficiency virus type 1 promoter. *J. Virol.* **69**:7437–7444.

[9] Blazevic, V., A. Ranki, S. Mattinen, S. L. Valle, S. Koskimies, G. Jung, and K. J. Krohn. 1993. Helper T-cell recognition of HIV-1 Tat synthetic peptides. *J. Acquir. Immune. Defic. Syndr.* **6**:881–890.

[10] Braddock, M., A. Thorburn, A. Chambers, A. Kingsman, and S. Kingman. 1990. A nuclear translational block imposed by the HIV-1 U3 region is relieved by the Tat-TAR interaction. *Cell* **62**:1123–1133.

[11] Brake, D. A., C. Debouk, and G. Biesecker. 1990. Identification of an Arg-Gly-Asp (RGD) cell adhesion site in human imuunodeficiency virus type 1 transactivation protein tat. *J. Cell Biol.* **111**:1275–1281.

[12] Buratowski, S. 1994. The basics of basal transcription by RNA polymerase II. *Cell* **77**:1–3.

[13] Buonaguro, L., F. Buonaguro, G. Giraldo, and B. Ensoli. 1994. The human immunodeficiency virus type 1 tat protein transactivates tumor necrosis factor beta gene expression through a TAR-like structure. *J. Virol.* **68**:2677–2682.

[14] Calnan, B. J., B. Tidor, S. Biancalana, D. Hudson, and A. D. Frankel. 1991. Arginine-mediated RNA recognition: the arginine fork. *Science* **252**:1167–1171.

[15] Carroll, R., L. Martarano, and D. Derse. 1991. Identification of lentivirus Tat functional domains through generation of equine infectious anemia virus/human immunodeficiency virus type 1 tat gene chimera. *J. Virol.* **65**:3460–3467.

[16] Chang, H. K., R. C. Gallo, and B. Ensoli. 1995. Regulation of cellular gene expression and function by the human immunodeficiency virus type 1 Tat protein. *J. Biomed. Sci* **2**:189–202.

[17] Chang, Y. N., and K. T. Jeang. 1992. The basic RNA-binding domain of HIV-2 Tat contributes to preferential trans-activation of a TAR2-containing LTR. *Nucleic Acids Res.* **20**:5465–5472.

[18] Chun, R., and K.-T. Jeang. 1996. Requirements for RNA polymeraseII carboxyl-terminal domain for activated transcription of human retroviruses human T-cell lymphotropic virus I and HIV-1. *J. Biol. Chem.* **271**:27888–27894.

[19] Churcher, M., C. Lamont, F. Hamy, C. Dingwall, S. Green, A. Lowe, J. Butler, M. Gait, and J. Karn. 1993. High affinity binding of TAR RNA by the human immunodeficiency virus type 1 tat protein requires base-pairs in the RNA stem and amino acids residues flanking the basic region. *J. Mol. Biol.* **230**:90–110.

[20] Cordingley, M. G., R. L. LaFemina, P. L. Callahan, J. H. Condra, V. V. Sardana, D. J. Graham, T. M. Nguyen, K. LeGrow, L. Gotlib, A. J. Schlabach, and R. J. Colonno. 1990. Sequence-specific interaction of Tat protein and Tat peptides with the transactivation-responsive sequence element of human immunodeficiency virus type 1 in vitro. *Proc. Natl. Acad. Sci. USA* **87**:8985–8989.

[21] Cullen, B. R. 1986. Trans-activation of human immunodeficiency virus occurs via a bimodal mechanism. *Cell* **46**:973–982.

[22] Cullen, B. R. 1993. Does HIV-1 Tat induce a change in viral initiation rights? *Cell* **73**:417.

[23] Dayton, A., J. Sodroski, C. Rosen, W. Goh, and W. Haseltine. 1986. The transactivator gene of the human T cell lymphotropic virus type III is required for replication. *Cell* **44**:941–947.

[24] Derse, D., M. Carvalho, R. Carroll, and B. M. Peterlin. 1991. A minimal lentivirus tat. *J. Virol.* **65**:7012–7015.

[25] Dingwall, C., I. Ernberg, M. J. Gait, S. M. Green, S. Heaphy, J. Karn, A. D. Lowe, M. Singh, M. A. Skinner, and R. Vallerio. 1989. Human immunodeficiency virus 1 tat protein binds trans-activation responsive region (TAR) RNA in vitro. *Proc. Natl. Acad. Sci. USA* **86**:6925–6929.

[26] Edery, I., R. Petryshyn, and N. Sonenberg. 1989. Activation of double-stranded RNA-dependent kinase (dsI) by the TAR region of HIV-1 mRNA: a novel translational control mechanism. *Cell* **56**:303–312.

[27] Ensoli, B., G. Barillari, S. Z. Salahuddin, R. C. Gallo, and F. Wong-Staal. 1990. Tat protein of HIV-1 stimulates growth of cells derived from Kaposi's sarcoma lesions of AIDS patients. *Nature* **345**:84–86.

[28] Ensoli, B., L. Buonaguro, G. Barillari, V. Fiuorelli, R. Gendelman, R. A. Morgan, P. Wingfield, and R. C. Gallo. 1993. Release, uptake, and effects of extracellular human immunodeficiency virus type 1 tat protein on cell growth and viral transactivation. *J. Virol.* **67**:277–287.

[29] Fisher, A., M. Feinberg, S. Josephs, M. Harper, L. Marselle, G. Reyes, M. Gonda, A. Aldovini, C. Debouk, R. C. Gallo, and F. Wong-staal. 1986. The transactivator gene of HTLV-III is essential for virus replication. *Nature* **320**:367–371.

[30] Frankel, A. D., D. S. Bredt, and C. O. Pabo. 1988. Tat protein from human immunodeficiency virus forms a metal-linked dimer. *Science* **240**:70–73.

[31] Garcia, J. A., D. Harrich, L. Pearson, R. Misuyasu, and R. Gaynor. 1988. Functional domains required for tat-induced transcriptional activation of the HIV-1 long terminal repeat. *EMBO J.* **7**:3143–3147.

[32] Gatignol, A., C. Buckler, and K. T. Jeang. 1993. Relatedness of an RNA-binding motif in human immunodeficiency virus type 1 TAR RNA-binding protein TRBP to human P1/dsI kinase and Drosophila staufen. *Mol. Cell. Biol.* **13**:2193–2202.

[33] Gatignol, A., A. Buckler-White, B. Berkhout, and K.-T. Jeang. 1991. Characterization of a human TAR RNA-binding protein that activates the HIV-1 LTR. *Science* **251**:1597–1600.

[34] Gatignol, A., M. Duarte, L. Daviet, Y. N. Chang, and K. T. Jeang. 1996. Sequential steps in Tat trans-activation of HIV-1 mediated through cellular DNA, RNA, and protein binding factors. *Gene Exp.* **5**:217–228.

[35] Gaynor, R., E. Soultanakis, M. Kuwabara, J. Garcia, and D. S. Sigman. 1989. Specific binding of a HeLa cell nuclear protein to RNA sequences in the human immunodeficiency virus transactivating region. *Proc. Natl. Acad. Sci. USA* **86**:4858–4862.

[36] Gibellini, D., A. Caputo, C. Celeghini, A. Bassini, P. M. La, S. Capitani, and G. Zauli. 1995. Tat-expressing Jurkat cells show an increased resistance to different apoptotic stimuli, including acute human immunodeficiency virus-type 1 (HIV-1) infection. *Br. J. Haematol.* **89**:24–33.

[37] Goldstein, G. 1996. HIV-1 Tat protein as a potential AIDS vaccine. *Nat. Med.* **1**:960–964.

[38] Gregoire, C. J., and E. P. Loret. 1996. Conformational heterogeneity in two regions of Tat results in structural variations of this protein as a function of HIV-1 isolates. *J. Biol. Chem.* **271**:22641–22646.

[39] Hauber, J., M. H. Malim, and B. R. Cullen. 1989. Mutational analysis of the conserved basic domain of human immunodeficiency virus tat protein. *J. Virol.* **63**:1181–1187.

[40] Herrmann, C. H., and A. P. Rice. 1995. Lentivirus Tat proteins specifically associate with a cellular protein kinase, TAK, that hyperphosphorylates the carboxyl-terminal domain of the large subunit of RNA polymerase II: candidate for a Tat cofactor. *J. Virol.* **69**:1612–1620.

[41] Howcroft, T. K., K. Strebel, M. A. Martin, and D. S. Singer. 1993. Repression of MHC class 2 gene promoter activity by 2 exon Tat of HIV. *Science* **260**:1320–1323.

[42] Huang, L. M., and K. T. Jeang. 1993. Increased spacing between Sp1 and TATAA renders human immunodeficiency virus type 1 replication defective: implication for Tat function. *J. Virol.* **67**:6937–6944.

[43] Huang, L. M., A. Joshi, R. Willey, J. Orenstein, and K. T. Jeang. 1994. Human immunodeficiency viruses regulated by alternative trans-activators: genetic evidence for a novel non-transcriptional function of Tat in virion infectivity. *EMBO J.* **13**:2886–2896.

[44] Jeang, K.-T., and B. Berkhout. 1992. Kinetics of HIV-1 LTR trans-activation: use of intragenic ribozyme to assess rate limiting steps. *J. Biol. Chem.* **267**:17891–17899.

[45] Jeang, K.-T., B. Berkhout, and B. Dropulic. 1993. Effects of integration and replication on the transcription of the HIV-1 LTR. *J. Biol. Chem.* **268**:24940–24949.

[46] Jeang, K.-T., Y. N. Chang, B. Berkhout, M.-L. Hammarskjold, and D. Rekosh. 1991. Regulation of HIV expression : mechanisms of action of Tat and Rev. AIDS 1991. A year in Review. *J. AIDS* 5:S3–S14.

[47] Jeang, K. T., R. Chun, N. H. Lin, A. Gatignol, C. G. Glabe, and H. Fan. 1993. In vitro and in vivo binding of human immunodeficiency virus type 1 Tat protein and Sp1 transcription factor. *J. Virol.* **67**:6224–6233.

[48] Jeang, K.-T., and A. Gatignol. 1994. Comparisons of regulatory features among primate lentiviruses. *Current Topics in Microbiology and Immunology*, in press.

[49] Jones, K. A., J. T. Kadonaga, P. A. Luciw, and R. Tjian. 1986. Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. *Science* **232**:755–759.

[50] Kao, S. Y., A. F. Calman, P. A. Luciw, and B. M. Peterlin. 1987. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature* **330**:489–493.

[51] Kashanchi, F., G. Piras, M. F. Radonovich, J. F. Duvall, R. Roeder, and J. N. Brady. 1994. Direct interaction of human TFIID with the HIV-1 transactivator tat. *Nature* **367**:295–299.

[52] Koken, S. E., A. E. Greijer, K. Verhoef, J. vanWamel, and B. Berkhout. 1994. Intracellular analysis of in vitro modified HIV tat protein. *J. Biol. Chem.* **269**:8366–8375.

[53] Kolson, D., J. Buchhalter, R. Collman, B. Hellmig, C. F. Farrell, C. Debouk, and F. Gonzalea-Scarano. 1993. HIV-1 tat alters normal organization of neurons and astrocytes in primary rodent brain cell cultures:RGD sequence dependence. AIDS Res. and Human Retroviruses **9**:677–685.

[54] Krone, W. J., C. Debouck, L. G. Epstein, P. Heutink, R. Meloen, and J. Goudsmit. 1988. Natural antibodies to HIV-tat epitopes and expression of HIV-1 genes in vivo. *J. Med. Virol.* **26**:261–270.

[55] Kuppuswamy, M., T. Subramanian, A. Srinivasan, and G. Chinnadurai. 1989. Multiple functional domains of Tat, the trans-activator of HIV-1, defined by mutational analysis. Nuc. Acids Res. **17**:3551–3561.

[56] Lamhamedi-Cherradi, S., B. Culmann-Penciolelli, B. Guy, M. P. Kieny, F. Dreyfus, A. G. Saimot, D. Sereni, D. Sicard, J.-P. Levy, and E. Gomard. 1992. Qualitative and quantitative analysis of human cytotoxic T-lymphocyte responses to HIV-1 proteins. *AIDS* **6**:1249–1258.

[57] Laspia, M. F., A. P. Rice, and M. B. Matthews. 1989. HIV-1 Tat protein increases transcriptional initiation and stabilizes elongation. *Cell* **59**:283–292.

[58] Li, C: J., D. J. Friedman, C. Wang, V. Metelev, and A. B. Pardee. 1995. Induction of apoptosis in uninfected lymphocytes by HIV-1 Tat protein. *Science* **268**:429–431.

[59] Liu, Y., Z. Wang, and T. M. Rana. 1996. Visualizing a specific contact in the HIV-1 Tat protein fragment and trans-activation responsive region RNA complex by photocross-linking. *J. Biol. Chem.* **271**:10391–10396.

[60] Lotz, M., I. Clark-Lewis, and V. Ganu. 1994. HIV-1 transactivator protein Tat induces proliferation and TGF beta expression in human articular chondrocytes. *J. Cell Biol.* **124**:365–371.

[61] Luo, Y., S. J. Madore, T. G. Parslow, B. R. Cullen, and B. M. Peterlin. 1993. Functional analysis of interactions between Tat and the trans-activation response element of human immunodeficiency virus type 1 in cells. *J. Virol.* **67**:5617–5622.

[62] Marciniak, R. A., B. J. Calnan, A. D. Frankel, and P. A. Sharp. 1990. HIV-1 Tat protein trans-activates transcription in vitro. *Cell* **63**:791–802.

[63] McMillan, N. A., R. F. Chun, D. P. Siderovski, J. Galabru, W. M. Toone, C. E. Samuel, T. W. Mak, A. G. Hovanessian, K. T. Jeang, and B. R. Williams. 1995. HIV-1 Tat directly interacts with the interferon-induced, double-stranded RNA-dependent kinase, PKR. *Virology* **213**:413–424.

[64] McPhee, D. A., B. E. Kemp, S. Cumming, D. Stapleton, I. D. Gust, and R. R. Doherty. 1988. Recognition of envelope and tat protein synthetic peptide analogs by HIV positive sera or plasma. *FEBS. Lett.* 1988. Jun. **233**:393–396.

[65] Metzger, A. U., T. Schindler, D. Willbold, M. Kraft, C. Steegborn, A. Volkmann, R. W. Frank, and P. Rosch. 1996. Structural rearrangements on HIV-1 Tat (32–72) TAR complex formation. *FEBS. Lett.* **384**:255–259.

[66] Meyerhans, A., R. Cheynier, J. Albert, M. Seth, S. Kwok, J. Sninsky, L. Morfeldt-Manson, B. Asjo, and S. Wain-Hobson. 1989. Temporal fluctuation in HIV quasispecies in vivo are not reflected by sequential HIV isolations. *Cell* **58**:901–910.

[67] Milani, D., M. Mazzoni, P. Borgatti, G. Zauli, L. Cantley, and S. Capitani. 1996 Extracellular human immunodeficiency virus type 1 Tat protein activates phosphatidylinositol 3-kinase in PC12 neuronal cells. *J. Biol. Chem.* **271**:22961–22964.

[68] Muesing, M., D. Smith, and D. Capon. 1987. Regulation of mRNA accumulation by human immunodeficiency virus trans-activator protein. *Cell* **48**:691–701.

[69] Nabell, L. M., R. H. Raja, P. P. Sayeski, A. J. Paterson, and J. E. Kudlow. 1994 Human immunodeficiency virus 1 Tat stimulates transcription of the transforming growth factor alpha gene in an epidermal growth factor-dependent manner. *Cell. Growth. Differ.* **5**:87–93.

[70] Neuveut, C., and K.-T. Jeang. 1996. Recombinant human immunodeficiency virus type 1 genomes with tat unconstrained by overlapping reading frames reveal residues in Tat important for replication in tissue culture. *J. Virol.* **70**:5572–5581.

[71] Niikura, M., G. Dornadula, H. Zhang, M. Mukhtar, D. Lingxun, K. Khalili, O. Bagasra, and R. J. Pomerantz. 1996. Mechanisms of transcriptional transactivation and restriction of human immunodeficiency virus type I replication in an astrocytic glial cell. *Oncogene.* **13**:313–322.

[72] Parkin, N. T., E. Cohen, A. Darveau, C. Rosen, W. Haseltine, and N. Sonenberg. 1988 Muational analysis of the 5' noncoding region of human immunodeficiency virus type 1: effects of secondary structure on translation. *EMBO J.* **7**:2831–2837.

[73] Peterlin, B. M., P. A. Luciw, P. J. Barr, and M. D. Walker. 1986. Elevated levels of mRNA can account for the trans-activation of human immunodeficiency virus. *Proc. Natl. Acad. Sci. USA* **83**:9734–9738.

[74] Purvis, S. F., D. L. Georges, T. M. Williams, and M. M. Lederman. 1992. Suppression of interleukin-2 and interleukin-2 receptor expression in Jurkat cells stably expressing the human immunodeficiency virus Tat protein. *Cell. Immunol.* **144**:32–42.

[75] Rappaport, J., S. J. Lee, K. Khalili, and F. Wong-Staal. 1989. The acidic amino-terminal region of HIV-1 tat protein constitutes an essential activating domain. *New Biologist* **1**:101–110.

[76] Rautonen, J., and N. Rautonen. 1992. Tat and Kawasaki disease. *Immunology Today* **13**:190–191.

[77] Re, M. C. 1995. Effect of antibody to HIV-1 Tat protein on viral replication in vitro and progression of HIV-1 disease in vivo. *J. AIDS* **10**:408–416.

[78] Reiss, P., F. Wolfe, C. L. Kuiken, A. deRonde, J. Dekker, C. A. B. Boucher, C. Debouck, J. M. A. Lange, and J. Goudsmit. 1991. Contribution of antibody response to recombinant HIV-1 gene-encoded products nef, rev, tat, and protease in predicting development of AIDS in HIV-1 infected individuals. *J. AIDS* **4**:165–172.

[79] Rice, A. P., and F. Carlotti. 1990. Mutational analysis of the conserved cystein-rich region of the human immunodeficiency virus type 1 tat protein. *J. Virol.* **64**:1864–1868.

[80] Rice, A. P., and F. Carlotti. 1990. Structural analysis of wild-type and mutant human immunodeficiency virus type 1 Tat proteins. *J. Virol.* **64**:6018–6026.

[81] Rice, A. P., and F. Chan. 1991. Tat protein of human immuondeficiency virus type 1 is a monomer when expressed in mammalian cells. *Virology* **185**:451–454.

[82] Rice, A., and M. B. Mathews. 1988. Transcriptional but not translational regulation of HIV-1 by the tat gene product. *Nature* **322**:551–553.

[83] Rodman, T. C., F. H. Pruslin, S. E. To, and R. Winston. 1992. Human immunodeficiency virus (HIV) Tat-reactive antibodies present in normal HIV-negative sera and depleted in HIV-positive sera. Identification of the epitope. *J. Exp. Med.* **175**:1247–1253.

[84] Rodman, T. C., S. E. To, H. Hashish, and K. Manchester. 1993. Epitopes for natural antibodies of human immunodeficiency virus (HIV)-negative (normal) and HIV-positive sera are coincident with two key functional sequences of HIV Tat protein. *Proc. Natl. Acad. Sci. USA* **90**:7719–7723.

[85] Rosch, P., and D. Willbold. 1996. Is EIAV Tat protein a homeodomain? *Science* **272**:1672.

[86] Rosen, C. A., J. G. Sodroski, W. C. Goh, A. I. Dayton, J. Lippke, and W. A. Haseltine. 1986. Post-transcriptional regulation accounts for the trans-activation of the human T-lymphotropic virus type III. *Nature* **319**:555–559.

[87] Roy, S., U. Delling, C.-H. Chen, C. A. Rosen, and N. Sonenberg. 1990. A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. *Genes Dev.* **4**:1365–1373.

[88] Roy, S., M. Katze, N. Parkin, I. Edery, and N. Sonenberg. 1990. Control of the interferon-induced 68-kilodalton protein kinase by the HIV-1 tat gene product. *Science* **247**:1216–1219.

[89] Roy, S., N. T. Parkin, C. Rosen, J. Itovitch, and N. Sonenberg. 1990. Structural requirements for trans-activation of human immunodeficiency virus type 1 long terminal repeat-directed gene expression by Tat: importance of base pairing, loop sequence, and bulges in the Tat-responsive sequence. *J. Virol.* **64**:1402–1406.

[90] Ruben, S., A. Perkins, R. Purcell, K. Joung, R. Sia, R. Burghoff, W. A. Haseltine, and C. A. Rosen. 1989. Structural and functional characterization of human immunodeficiency virus tat protein. *J. Virol.* **63**:1–8.

[91] Sabatier, J., E. Vives, K. Mabrouk, A. Benjouad, H. Rochat, A. Duval, B. Hue, and E. Bahraoui. 1991. Evidence for neurotoxic activity of tat from human immunodeficiency virus type 1. *J. Virol.* **65**:961–967.

[92] Sadaie, M. R., J. Rappaport, T. Benter, S. F. Josephs, R. Willis, and F. Wong-Staal. 1988. Missense mutations in an infectious human immunodeficiency viral genome: functional mapping of tat and identification of the rev splice acceptor. *Proc. Natl. Acad. Sci. USA* **85**:9224–9228.

[93] Scala, G., M. R. Ruocco, C. Ambrosino, M. Mallardo, V. Giordano Baldassarre, F., E. Dragonetti, I. Quinto, and S. Venuta. 1994. The expression of the interleukin 6 gene is induced by the human immunodeficiency virus 1 tat protein. *J. Exp. Med.* **179**:961–971.

[94] Seigel, L. J., L. Ratner, S. F. Josephs, D. Derse, M. Feinberg, G. A. Reyes, S. J. O'Brien, and F. Wong-Staal. 1986. Trans-activation induced by human T-lymphotropic virus type III (HTLV-III) maps to a viral sequence encoding 58 amino acids and lacks tissue specificity. *Virology* **148**:226–231.

[95] Selby, M. J., and B. M. Peterlin. 1990. Trans-activation by HIV-1 Tat via a heterologous RNA binding protein. *Cell* **62**:769–776.

[96] SenGupta, D. N., B. Berkhout, A. Gatignol, A. Zhou, and R. H. Silverman. 1990. Direct evidence for translation regulation by leader RNA and Tat protein in human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. USA* **87**:7492–7496.

[97] SenGupta, D. N., and R. H. Silverman. 1989. Activation of interferon-regulated, dsRNA-dependent enzymes by HIV-1 leader RNA. *Nucl. Acids. Res.* **17**:969–978.

[98] Sheline, C. T., L. H. Milocco, and K. A. Jones. 1991. Two distinct nuclear transcription factors recognize loop and bulge residues of the HIV-1 TAR RNA hairpin. *Genes Dev.* **5**:2508–2520.

[99] Siderovski, D. P., T. Matsuyama, E. Frigerio, S. Chui, X. Min, H. Erfle, M. Sumner-Smith, R. W. Barnett, and T. W. Mak. 1992. Random mutagenesis of the human immunodeficiency virus type-1 trans-activator of transcription (HIV-1 Tat). *Nuc. Acids Res.* **20**:5311–5320.

[100] Southgate, C. D., and M. R. Green. 1995. Delineating minimal protein domains and promoter elements for transcriptional activation by lentivirus Tat proteins. *J. Virol.* **69**:2605–2610.

[101] Southgate, C., M. L. Zapp, and M. R. Green. 1990. Activation of transcription by HIV-1 Tat protein tethered to nascent RNA through another protein. Nature **345**:640–642.

[102] Steinaa, L., A. M. Sorenson, J. O. Nielsen, and J.-E. S. Hansen. 1994. Tat inhibits replication of virus in culture. *Arch. Virol.* **139**:263–271.

[103] Tiley, L. S., P. H. Brown, and B. R. Cullen. 1990. Does the human immunodeficiency virus tat transactivator contain a discrete activation domain? *Virology* **178**:560–567.

[104] Tong-Starksen, S., A. Baur, X. B. Lu, E. Peck, and B. M. Peterlin. 1993. Second exon of Tat of HIV-2 is required for optimal transactivation of HIV-1 and HIV-2 LTRs. *Virology* **195**:826–830.

[105] Viglianti, G., and J. I. Mullins. 1988. Functional comparison of transactivation by simian immunodeficiency virus from rhesus macaques and human immunodeficiency virus type 1. *J. Virol.* **62**:4523–4532.

[106] Wang, Z., and T. M. Rana. 1996. RNA conformation in the Tat-TAR complex determined by site-specific photo-cross-linking. *Biochemistry* **35**:6491–6499.

[107] Wang, Z., X. Wang, and T. M. Rana. 1996. Protein orientation in the Tat-TAR complex determined by psoralen photo-crosslinking. *J. Biol. Chem.* **271**:16995–16998.

[108] Weeks, K. M., and D. M. Crothers. 1991. RNA recognition by Tat-derived peptides: interaction in the major groove? *Cell* **66**:577–588.

[109] Wu, F., J. Garcia, D. Sigman, and R. Gaynor. 1991. Tat regulates binding of the human immunodeficiency virus trans-activating region RNA loop-binding protein TRP-185. *Genes Dev.* **5**:2128–2140.

[110] Zawel, L, and D. Reinberg. 1995. Common themes in assembly and function of eukaryotic transcription complexes. *Annu. Rev. Biochem* **64**: 533–561.

[111] Zhou, Q., and P. A. Sharp. 1996. Tat-SF1: Cofactor for stimulation of transcriptional elongation by HIV-1 Tat. *Science* **274**:605–609.

Structure and Function of HIV-1 Vpu

Klaus Strebel

LMM, NIAID, National Institutes of Health, Bethesda, MD 20892–0460

Introduction

Aside from the typical retroviral *gag*, *pol*, and *env* genes, HIV encodes a series of accessory genes, *vif*, *vpr*, *vpx*, *vpu*, and *nef*. The *vpu* gene is found exclusively in HIV-1 (Strebel et al., 1988; Cohen et al, 1988; Matsuda, 1988). In tissue culture systems, defects in accessory genes are frequently not correlated with a detectable impairment of virus replication and, as a result, these genes are often referred to as "non-essential". However, as more and more information on the function of the accessory proteins becomes available it also becomes increasingly clear that *in vivo*, these proteins indeed exert important functions. With respect to Vpu, Li et al found that macaques infected with Vpu-negative simian-human immunodeficiency virus chimeras (SHIV) had lower virus loads than Vpu-positive virus (Li et al., 1995). Also, using a SCID-hu model, Aldrovandi & Zack demonstrated that deletion of Vpu significantly affects virus infectivity and, to a somewhat lesser extend, pathogenicity of HIV (Aldrovandi & Zack, 1996). Thus, while there is currently only limited information available on the importance of Vpu *in vivo*, such model systems might be useful in investigating the *in vivo* relevance of Vpu. This article attempts to provide a brief overview on our current understanding of this viral accessory factor. More exhaustive reviews on Vpu can be found elsewhere (Jabbar, 1995; Bour et al., 1995a; Trono, 1995).

Structural Considerations

Vpu is a small integral transmembrane protein which is cotranslationally inserted into membranes of infected cells (Strebel et al., 1989). The protein consists of an N-terminal hydrophobic domain, which functions both as signal peptide and membrane anchor, and a hydrophilic C-terminal domain which protrudes into the cytoplasm. Vpu contains two highly conserved seryl residues, located in the cytoplasmic domain, which are phosphorylated by the ubiquitous protein kinase CK-2 (Schubert, et al., 1992; Schubert et al., 1994; Friborg et al., 1995). Phosphorylation is essential for at least one of the biological functions of Vpu as discussed below. Vpu forms homo-oligomeric structures (Maldarelli et al., 1993), a feature that may be critical for a proposed ion channel function of Vpu (see below). Based on 2D 1H NMR spectroscopy of a peptide corresponding to the cytoplasmic domain of Vpu (Wray et al., 1995; Federau et al., 1996), it was proposed that the cytoplasmic domain of Vpu contains two α-helical domains, helix-1 and helix-2 (Figure 1), which are connected by an unstructured region containing the two conserved phosphoseryl residues. In addition, computer models predict a third α-helical domain in the transmembrane domain of Vpu, which could play an important role in the formation of ion channels (Schubert et al., 1996b).

The *vpu* gene overlaps at its 3′-end with the *env* gene (Figure 2). Indeed, Vpu and Env are expressed from the same bicistronic mRNA in a Rev-dependent manner (Schwartz et al., 1990), presumably by leaky scanning of ribosomes through the *vpu* initiation codon. In tissue culture, this arrangement results in the synthesis of roughly equimolar levels of Vpu and Env proteins and it is possible that this unusual utilization of viral transcripts might reflect a requirement for the coordinate action of the two viral gene products. Several HIV-1 isolates were found to carry point mutations in the Vpu translation initiation codon but have otherwise intact *vpu* genes. Since removal of the Vpu initiation codon results in increased expression of the downstream *env* gene, it is possible that HIV-1 actually uses this mechanism as a molecular switch to regulate the relative expression of Vpu or Env in infected cells. The possible benefits of such a regulation are unclear; however, it is conceivable that, under certain circumstances, *vpu*-defective isolates expressing increased levels of Env protein have a selective advantage over "wild-type" viruses expressing Vpu.

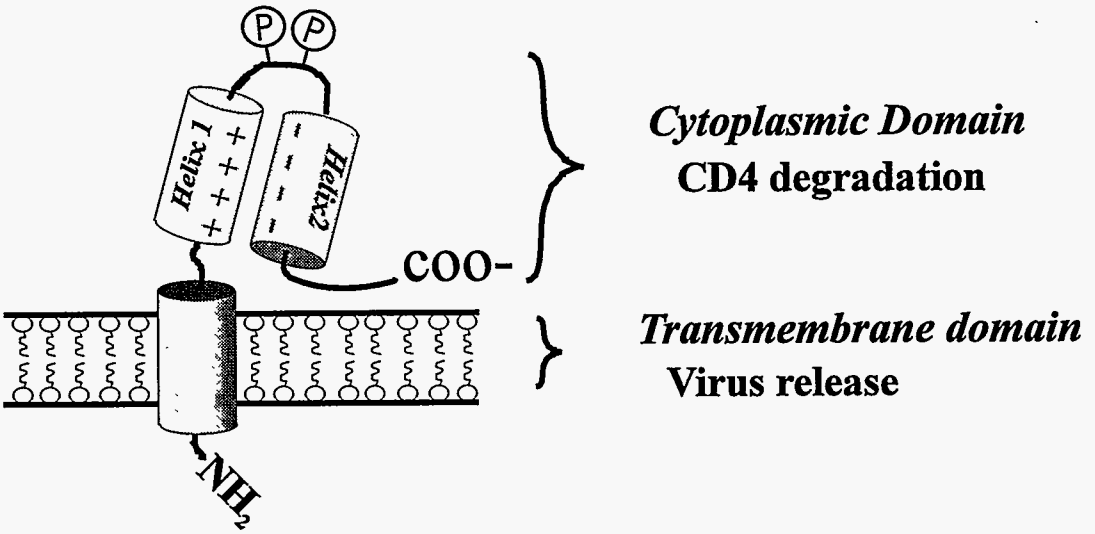


Figure 1. Structural domains of Vpu. Vpu consists of an N-terminal hydrophobic domain, that functions as membrane anchor, and a hydrophilic cytoplasmic domain. The cytoplasmic domain contains two amphipathic α-helical domains of opposite polarity. They are separated by an unstructured region containing two conserved seryl residues which are phosphorylated by protein kinase CK-2. The cytoplasmic domain contains seqences critical for CD4 degradation while the membrane anchor domain has a critical function in regulating virus release and plays an important role in the formation of cation selective ion channels. Vpu forms homo-oligomeric complexes. Only the monomeric form is shown. A putative interaction between helix 1 and helix-2 as shown in the cartoon suggests only one of many possible conformations of Vpu.

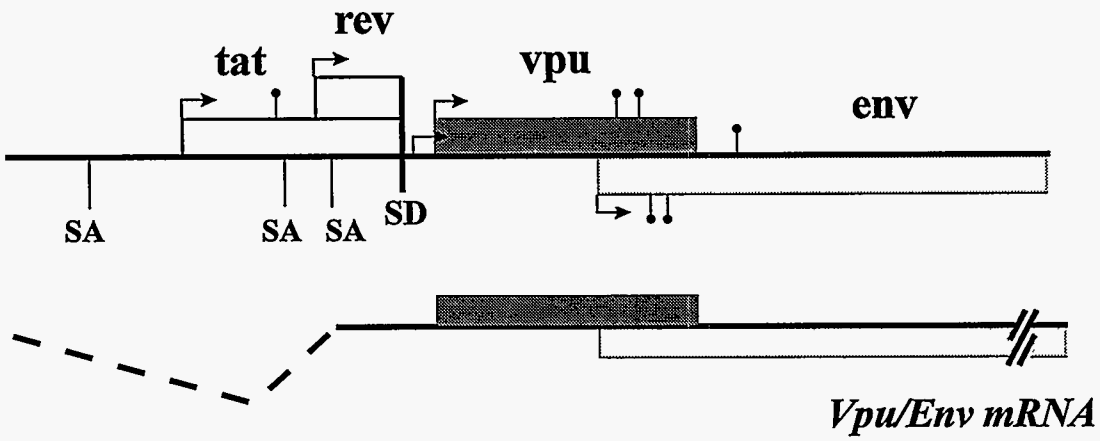


Figure 2. Structure of the *in vpu* gene. The *vpu* gene overlaps at its 3′-end with the *env* gene. Both Vpu and Env are expressed from the same bicistronic mRNA at roughly equimolar levels. The positions of AUG codons at the beginning of open reading frames are marked by arrows. Internal AUG codons are denoted by filled circles. SA = splice acceptor sites; SD = splice donor site.

Vpu has two primary biological activities which are discussed in detail below and are summarized in Figure 3. These include the degradation of CD4 in the endoplasmic reticulum and the augmentation of virus secretion from the plasma membrane. In addition, expression of Vpu has been associated with a reduction in syncytia formation of infected cells (Strebel et al., 1989; Terwilliger et al., 1989; Klimkait et al., 1990; Yao et al., 1993; Schubert et al., 1994). This latter phenomenon may be a consequence of the reduced presence of viral Env protein at the cell surface (Yao et al., 1993) due to the more efficient shedding of viral particles in the presence of Vpu. Aside from that, several other, less well defined functions have been associated with Vpu activity. These include the regulation of ER-to-Golgi transport of proteins (Vincent & Jabbar, 1995), or the modulation of MHC I antigen complexes (Kerkau, 1995).

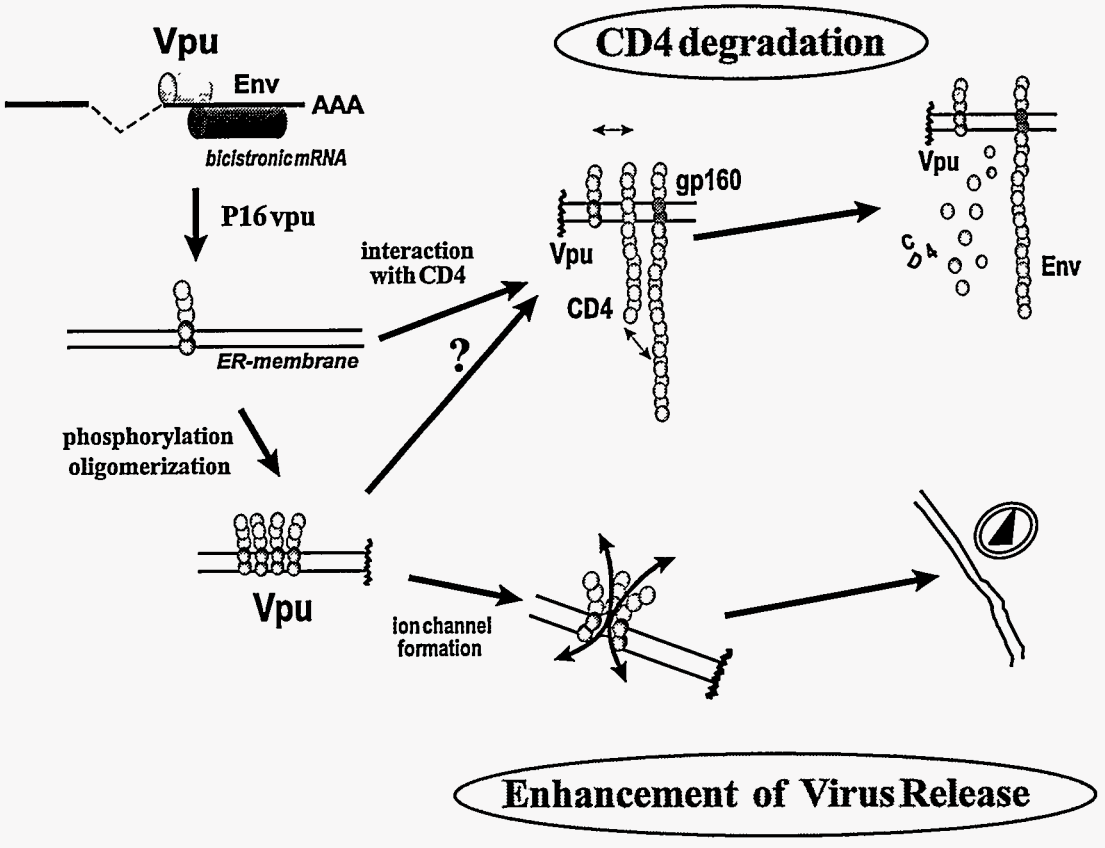


Figure 3. Vpu enhances virus particle secretion and induces degradation of CD4. The 81 amino acid Vpu protein (P16vpu) is synthesized from a bicistronic mRNA and cotranslationally inserted into membranes of the endoplasmic reticulum. Vpu is phosphorylated by CK-2 and forms homo- oligomeric structures. Newly synthesized CD4 is retained in the ER due to the formation of stable complexes with Env protein. Vpu physically associates with CD4 thereby triggering a mechanism that results in the destruction of CD4 and in the release of Env protein from the ER blockage. It is unclear whether this function of Vpu involves monomeric or oligomeric Vpu (? in the cartoon). Vpu exits the ER and accumulates at or near the Golgi. In cells overexpressing Vpu, small amounts of Vpu reach the plasma membrane. Oligomeric forms of Vpu have the capacity to form ion channels which presumably are involved in the regulation of virus particle release.

Vpu Facilitates the Release of Virus Particles

The original biological phenotype associated with viruses lacking a functional *vpu* gene was the impairment of virus particle secretion from infected cells (Strebel et al., 1988; Terwilliger et al.,

1989). This defect is manifested by the increased budding of viruses from internal membranes and the accumulation of budding particles at the cell surface where they remain loosely attached to the plasma membrane (Klimkait et al., 1990). Such particles can be released however, by vigorously shaking the cultures, and they are fully infectious (Klimkait et al., 1990). This suggests that Vpu regulates one of the final steps in virus production that is required for the efficient detachment of virions from the plasma membrane. Although it has been generally accepted that Vpu augments virus secretion from a variety of human cells, including PBMC, macrophages or CD4+ T cell lines, as well as non-T cell lines such as epithelial HeLa cells or SW480 cells (Strebel et al., 1988; Terwilliger et al., 1989; Klimkait et al., 1990; Westervelt et al., 1992; Yao et al., 1992; Göttlinger et al., 1993; Balliet et al., 1994; Kawamura et al., 1994; Schubert et al., 1994; Schubert et al., 1995; Sakai et al., 1995; Theodore et al., 1996), the enhancing effect of Vpu varies between different cell types and has been reported to range from a mere 2-to 3-fold in PBLs to up to 1000 fold in primary macrophages. Despite these cell type specific variations, the principal function of Vpu is not restricted to certain human cell types. In contrast, Geraghty et al. and Göttlinger et al. who analyzed the function of Vpu in Cos-1, Cos-7, or CV1 cells, were unable to detect any Vpu-mediated enhancement of virus secretion in simian cell lines (Geraghty et al., 1994; Göttlinger et al., 1991). It is possible that the efficient and Vpu-independent release of viruses from Cos cells is due to the expression of a cellular eqivalent to Vpu in simian cells.

The mechanistic details of Vpu activity are not well understood; however, several models can be proposed that are consistent with existing experimental data. First, Vpu could actively facilitate detachment of progeny virions from the cell surface, for example by altering the characteristics of the plasma membrane or by actively facilitating membrane fusion similar to the function exerted by annexin VI, which is required for budding of clathrin-coated pits during endocytosis (Lin et al., 1992). At least in the latter case, Vpu would be required in significant amounts at the cell surface. However, while Vpu was indeed observed at the plasma membrane of cells when overexpressed, the bulk of the protein is clearly sequestered on internal membranes (Maldarelli, personal communication; Schubert et al., 1996a). In addition, attempts to demonstrate the presence of Vpu in virions have thus far failed, also arguing against the presence of significant amounts of Vpu at the cell surface. Alternatively, Vpu could regulate virus release from internal membranes either by passively preventing budding on internal membranes, thus redirecting or restricting virus assembly to the cell surface, or by indirectly affecting particle release through the modulation of so far unidentified cellular factors.

Recent evidence suggests that Vpu has the ability to form cation selective ion channels (Ewart et al., 1996; Schubert et al., 1996b); there is some evidence that this activity of Vpu correlates with its ability to regulate virus release. The ability of Vpu to form ion channels requires the integrity of the TM domain, and alterations in the primary structure of the TM domain not only abolish the ability of Vpu to form ion conductive pores but also negate the capacity to regulate virus release (Schubert, 1996a/b). While these findings imply that regulation of virus release is mediated through an ion channel activity of Vpu, it remains to be shown how an ion channel activty of Vpu could affect the detachment of budding particles from the plasma membrane or how it could redirect virus budding from internal membranes to the plasma membrane.

Vpu Induces CD4 Degradation

A second function which has been extensively investigated is the ability of Vpu to induce degradation of CD4 at the endoplasmic reticulum (ER). One of the complications that HIV faces when replicating in CD4+ cells is the formation of stable complexes between cellular CD4 and the HIV Env protein. Such complexes are trapped in the ER (Bour et al., 1991; Crise et al., 1990; Kawamura et al., 1989; Jabbar & Nayak, 1990; Willey et al., 1992a/b) thereby preventing transport of both CD4 and Env to the cell surface. In the presence of Vpu, however, Env was found to be liberated from CD4/Env complexes (Willey et al., 1992a; Kimura et al., 1994) concomitant with a significant reduction in the detectable steady state levels of CD4. The reduction of CD4 was caused by Vpu-induced degradation which reduced the half-life of CD4 in Hela cells from normally 4–6 hours to approximately 10 minutes (Willey et al., 1992b). Efficient degradation of CD4 requires its retention in the ER. This is normally accomplished by the formation of stable complexes with HIV Env. However, Env protein per se is

not involved in CD4 degradation since it is not required when CD4 is artificially retained in the ER (Willey, 1992b). Also, CD4 degradation can be observed in an *in vitro* translation system where no viral proteins other than Vpu are required (Chen et al., 1993). Unlike Nef, Vpu is unable to target cell surface CD4, or, for that matter, CD4 that has exited the ER. Nevertheless, the rapid degradation of de novo synthesized CD4 in the presence of Vpu ultimately leads to a depletion of cell surface CD4 due to the reduction of the overall cellular CD4 pool.

Based on mutational analysis of CD4, it is known that deletion of the cytoplasmic domain of CD4 renders the protein insensitive to Vpu, suggesting that this domain of CD4 contains Vpu-responsive sequences (Chen et al., 1993; Lenburg & Landau, 1993; Vincent et al., 1993). This is supported by the fact that transfer of CD4 cytoplasmic sequences to CD8, a cellular receptor molecule that is not normally targeted by Vpu, resulted in Vpu-dependent degradation of the chimeric molecules (Willey et al., 1994). In fact, transfer of a 18 amino acid membrane proximal fragment of the CD4 cytoplasmic domain (amino acids 403 to 420) into the CD8 cytoplasmic tail was sufficient to confer Vpu sensitivity. While these results suggest that cytoplasmic sequences of CD4 are sufficient to confer Vpu sensitivity to heterologous membrane proteins, other studies suggest that sequences located in the TM domain of CD4 may also be required (Raja et al., 1994; Buonocore et al., 1994). It is possible that structural constraints imposed by the extracellular or TM domains of the model proteins employed in the different studies explain the apparently differing results. While the importance of the CD4 TM domain for this process thus remains unclear, the importance of the CD4 cytoplasmic domain is undisputed. It is now clear that this domain of CD4, which is predicted to form an α-helical structure (Yao et al., 1995), is involved in the physical interaction with the cytoplasmic domain of Vpu (Bour et al., 1995b; Schubert et al., 1996a; Margottin et al., 1996). Such interaction between CD4 and Vpu appears to be a prerequisite for CD4 degradation but is in itself not sufficient to trigger CD4 degradation. This is evidenced by the fact that mutants of Vpu that are unable to induce CD4 degradation are still able to bind to CD4 (Bour et al., 1995b).

Mutational analysis of Vpu demonstrated that phosphorylation of Vpu at two conserved seryl residues, Ser52 and Ser56, in its cytoplasmic domain is essential for its ability to induce CD4 degradation (Schubert, 1994; Friborg et al., 1995) but not for its ability to bind to CD4 (Bour et al., 1995b). The Vpu TM domain is not required for the interaction with CD4 (Margottin et al., 1996) and alterations in the Vpu TM domain have no apparent effect on CD4 degradation, provided the protein retains its ability to properly associate with membranes (Schubert et al., 1996a; Friborg et al., 1994). These results suggest that sequences critical for the induction of CD4 degradation are located in the Vpu cytoplasmic domain while the TM domain merely serves as a membrane anchor. How Vpu triggers degradation of CD4 is still unclear. It is conceivable that binding of Vpu to CD4 induces a conformational change in CD4 which in turn activates a cellular pathway designed to eliminate aberrantly folded proteins. However, the fact that CD4-Vpu interaction is necessary but not sufficient to induce CD4 degradation indicates that Vpu performs a catalytic function beyond the binding step.

Functional domains of Vpu and Vpu-like activities in HIV-2

In light of the multiple biological activities provided by the Vpu protein during the HIV-1 life cycle, it is intriguing that, except for the chimpanzee isolate SIVcpz (Huet, et al, 1990), no functional equivalent to Vpu is found in related viruses such as HIV-2 and SIV. This fact seems especially paradoxical since Vpu was shown to augment the release of chimeric viruses bearing the gag-pol regions of retroviruses that naturally lack a *vpu* ORF, such as HIV-2, visna virus and Moloney murine leukemia virus (Göttlinger et al., 1993). By examining the efficiency of particle release and Vpu-responsiveness of the ROD10 full-length molecular clone of HIV-2, we recently showed that mutations that disrupt the pROD10 env ORF, but not the *vif*, *vpr*, *vpx* or *nef* ORFs, have a profound negative effect on virus particle release (Bour et al., 1996a). Concomitantly, the pROD10 envelope glycoprotein provided in *trans* could rescue the envelope mutants and restore wild-type levels of particle release (Bour et al., 1996a). Similar results were independently reported by Ritter et al. who used a different HIV-2 isolate, HIV-2/ST (Ritter et al., 1996). The efficiency with which the ROD10 Env protein enhanced HIV-2 particle release was very similar to that of Vpu. Both activities could be provided in *trans* and were sensitive to treatment

of cells with Brefeldin A, suggesting that they both operate in a post-ER compartment (Bour et al., 1996a). Like Vpu, HIV-2 Env can regulate release not only of HIV-2 but of HIV-1 or SIV particles as well (Bour et al, 1996b). When compared in parallel, the effects of Vpu or ROD10 Env on HIV-1 or SIV virus release were identical. These findings suggest that the HIV-2 Env protein has a true Vpu-like activity with respect to virus particle release. However, in contrast to Vpu, there is no evidence that HIV-2 Env can induce degradation of CD4 (Bour et al., 1996b). This suggests that the activity of Vpu on CD4 is HIV-1-specific, a notion that may shed some light on the evolutionary relationship between the HIV-2 *env* and *vpu* genes. The selection pressure that promoted the appearance of the *vpu* gene in HIV-1 as well as the unique activity of Vpu on CD4 degradation may have been the increased affinity of the HIV-1 envelope for the CD4 receptor (Hoxie et al., 1991; Ivey-Hoyle et al., 1991; Mulligan et al., 1992), leading to more stable intracellular complexes and trapping of the envelope glycoprotein in the endoplasmic reticulum. The fact that the HIV-2 Env has a Vpu-like activity on particle release but not on CD4 degradation suggests that the ancestral activity of Vpu is particle release while the activity on CD4 was developed specifically by HIV-1 to counteract increased affinity of HIV-1 Env with the CD4 receptor.

The domain in HIV-2 Env which is responsible for the regulation of virus release has yet to be determined. Ritter et al. reported that an isolate containing a full-length 164 amino acid cytoplasmic tail is capable of performing this function while a variant, carrying a truncated, 17 amino acid cytoplasmic domain was inactive. It is tempting to speculate that the length of the cytoplasmic tail regulates this function of the HIV-2 Env protein. However, the ROD10 isolate used in our own studies carries only a short, 18 amino acid cytoplasmic tail yet efficiently activates virus release (Bour et al., 1996b). Thus it is likely that sequences other than the cytoplasmic domain of HIV-2 Env play a crucial role.

References

Aldrovandi, G. M., and J. A. Zack. 1996. Replication and pathogenicity of human immunodeficiency virus type 1 accessory gene mutants in SCID-hu mice. *J. Virol.* **70**:1505–1511.

Balliet, J. W., D. L. Kolson, G. Eiger, F. M. Kim, K. A. McGann, A. Srinivasan, and R. Collman. 1994. Distinct effects in primary macrophages and lymphocytes of the human immunodeficiency virus type 1 accessory genes *vpr*, *vpu*, and *nef*: mutational analysis of a primary HIV-1 isolate. *Virology* **200**:623–631.

Bour, S., F. Boulerice, and M. A. Wainberg. 1991. Inhibition of gp160 and CD4 maturation in U937 cells after both defective and productive infections by human immunodeficiency virus type 1. *J. Virol.* **65**:6387–6396.

Bour, S., R. Geleziunas, and M. Wainberg. 1995a. The human immunodeficiency virus type 1 (HIV-1) CD4 receptor and its central role in promotion of HIV-1 infection. *Microbiol. Rev.* **59**:63–93.

Bour, S., U. Schubert, and K. Strebel. 1995b. The human immunodeficiency virus type 1 Vpu protein specifically binds to the cytoplasmic domain of CD4: Implications for the mechanism of degradation. *J. Virol.* **69**:1510–1520.

Bour, S., U. Schubert, K. Peden, and K. Strebel. 1996a. The envelope glycoprotein of human immuunodeficiency virus type 2 enhances particle release: a Vpu-like factor? *J. Virol.* **70**:820–829.

Bour, S., and K. Strebel. 1996b. The human immunodeficiency virus (HIV) type 2 envelope protein is a functional complement to HIV type 1 Vpu that enhances particle release of heterologous retroviruses. *J. Virol.* 70:in press.

Buonocore, L., T. G. Turi, B. Crise, and J. K. Rose. 1994. Stimulation of heterologous protein degradation by the Vpu protein of HIV-1 requires the transmembrane and cytoplasmic domains of CD4. *Virology* **204**:482–486.

Chen, M. Y., F. Maldarelli, M. K. Karczewski, R. L. Willey, and K. Strebel. 1993. Human immunodeficiency virus type 1 Vpu protein induces degradation of CD4 *in vitro*: the cytoplasmic domain of CD4 contributes to Vpu sensitivity. *J. Virol.* **67**:3877–3884.

Cohen, E. A., E. F. Terwilliger, J. G. Sordroski, and W. A. Haseltine. 1988. Identification of a protein encoded by the vpu gene of HIV-1. *Nature* **334**:532–534

Crise, B., L. Buonocore, and J. K. Rose. 1990. CD4 is retained in the endoplasmic reticulum by the human immunodeficiency virus type 1 glycoprotein precursor. *J. Virol.* **64**:5585–5593.

Ewart, G. D., T. Sutherland, P. W. Gage, and G. B. Cox. 1996. The Vpu protein of human immunodeficiency virus type 1 forms cation-selective ion channels. *J. Virol.* **70**:7108–7115.

Federau, T., U. Schubert, J. Flossdorf, P. Henklein, D. Schomburg, and V. Wray. 1996. Solution structure of the cytoplasmic domain of the human immunodeficiency virus type 1 encoded virus protein U (Vpu). *Int. J. Peptide Res.* **47**:297–310.

Friborg, J., X. J. Yao, F. Boisvert, S. Garzon, and E. A. Cohen. 1994. Mutational analysis of the HIV-1 Vpu protein. *Leukemia* **8**:S156–S162.

Friborg, J., A. Ladha, H. Göttlinger, W. A. Haseltine, and E. A. Cohen. 1995. Functional analysis of the phosphorylation sites on the human immunodeficiency virus type 1 Vpu protein. *J. Acquired Immune Def. Syndr. Hum Retrovir* **8**:10–22.

Geraghty, R. J., K. J. Talbot, M. Callahan, W. Harper, and A. T. Panganiban. 1994. Cell type-dependence for Vpu function. *J. Med Primatol.* **23**:146–150.

Göttlinger, H. G., T. Dorfman, J. G. Sodroski, and W. A. Haseltine. 1991. Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc. Natl. Acad. Sci. USA* **88**:3195–3199.

Göttlinger, H. G., T. Dorfman, E. A. Cohen, and W. A. Haseltine. 1993. Vpu protein of human immunodeficiency virus type 1 enhances the release of capsids produced by gag gene constructs of widely divergent retroviruses. *Proc. Natl. Acad. Sci. USA* **90**:7381–7385.

Hoxie, J. A., L. F. Brass, C. H. Pletcher, B. S. Haggarty, and B. H. Hahn. 1991. Cytopathic variants of an attenuated isolate of human immunodeficiency virus type 2 exhibit increased affinity for CD4. *J. Virol.* **65**:5096–5101.

Huet, T., R. Cheynier, A. Meyerhans, G. Roelants, and S. Wain-Hobson. 1990. Genetic organization of a chimpanzee lentivirus related to HIV-1. *Nature* **345**:356–359.

Ivey-Hoyle, M., J. S. Culp, M. A. Chaikin, B. D. Hellmig, T. J. Matthews, R. W. Sweet, and M. Rosenberg. 1991. Envelope glycoproteins from biologically diverse isolates of immunodeficiency viruses have widely different affinities for CD4. *Proc. Natl. Acad. Sci. USA* **88**:512–516.

Jabbar, M. A., and D. P. Nayak. 1990. Intracellular interaction of human immunodeficiency virus type 1 (ARV-2) envelope glycoprotein gp160 with CD4 blocks the movement and maturation of CD4 to the plasma membrane. *J. Virol.* **64**:6297–6304.

Jabbar, M. A., 1995. The human immunodeficiency virus type 1 Vpu protein: Roles in virus release and CD4 downregulation. *Curr Top Microbiol Immunol* **193**:107–120.

Kawamura, I., Y. Koga, N. Oh-Hori, K. Onodera, G. Kimura, and K. Nomoto. 1989. Depletion of the surface CD4 molecule by the envelope protein of human immmunodeficiency virus expressed in a human CD4+ monocytoid cell line. *J. Virol.* **63**:7348–3754.

Kawamura, M., T. Ishizaki, A. Ishimoto, T. Shioda, T. Kitamura, and A. Adachi. 1994. Growth ability of human immunodeficiency virus type 1 auxillary mutants in primary blood macrophage cultures. *J. Gen. Virol.* **75**:2427–2431.

Kerkau, T., U. Schubert, T. Hünig, and A. Schimpl. 1995. Human immunodeficiency virus type 1 Vpu protein contributes to the downregulation of MHC class I molecules on HIV-1 infected cells. *J. Cell. Biochem.* Suppl 21B: 192

Kimura, T., M. Nishikawa, and A. Ohyama. 1994. Intracellular membrane traffic of human immunodeficiency virus type 1 envelope glycoproteins: Vpu liberates golgi-targeted gp160 from CD4-dependent retention in the endoplasmic reticulum. *J. Biochem.* **115**:1010–1020.

Klimkait, T., K. Strebel, M. D. Hoggan, M. A. Martin, and J. M. Orenstein. 1990. The human immunodeficiency virus type 1-specific protein vpu is required for efficient virus maturation and release. *J. Virol.* **64**:621–629.

Lenburg, M. E., and N. R. Landau. 1993. Vpu-induced degradation of CD4: requirement for specific amino acid residues in the cytoplasmic domain of CD4. *J. Virol.* **67**:7238–7245.

Li, J. T., M. Halloran, C. I. Lord, A. Watson, J. Ranchalis, M. Fung, N. L. Letvin, and J. G. Sodroski. 1995. Persistent infection of macaques with simian-human immunodeficiency viruses. *J. Virol.* **69**:7061–7071.

Lin, H. C., T. C. Suedhof, and R. G. W. Anderson. 1992. Annexin VI is required for budding of clathrin-coated pits. *Cell* **70**:283–291.

Maldarelli, F., M. Y. Chen, R. L. Willey, and K. Strebel. 1993. Human immunodeficiency virus type 1 Vpu protein is an oligomeric type 1 integral membrane protein. *J. Virol.* **67**:5056–5061.

Margottin, F., S. Benichou, H. Durand, V. Richard, L. X. Liu, E. Gomas, and R. Benarous. 1996. Interaction between the cytoplasmic domains of HIV-1 Vpu and CD4: role of Vpu residues involved in CD4 interaction and *in vitro* CD4 degradation. *Virology* **223**:381–386.

Matsuda, Z., M. J. Chou, M. Matsuda, J. H. Huang, Y. M. Chen, R. Redfield, K. Mayer, M. Essex, and T. H. Lee. 1988. Human immunodeficiency virus type 1 has an additional coding sequence in the central region of the genome. Proc. Natl. Acad. Sci. USA **85**:6968–6972.

Mulligan, M. J.; G. V. Yamshchikov, G. Ritter, Jr., F. Gao, M. J. Jin, C. D. Nail, C. P. Spies, B. H. Hahn, and R. W. Compans. 1992. Cytoplasmic domain truncation enhances fusion activity by the exterior glycoprotein complex of human immunodeficiency virus type 2 in selected cell types. *J. Virol.* **66**:3971–3975.

Raja, N. U., M. J. Vincent, and M. A. Jabbar. 1994. Vpu-mediated proteolysis of gp160/CD4 chimeric envelope glycoproteins in the endoplasmic reticulum: requirement of both the anchor and cytoplasmic domains of CD4. *Virology* **204**:357–366.

Ritter, G. D., G. Yamshchikov, S. J. Cohen, and M. J. Mulligan. 1996. Human immunodeficiency virus type 2 glycoprotein enhancement of particle budding: role of the cytoplasmic domain. *J. Virol.* **70**:2669–2673.

Sakai, H., K. Tokunaga, M. Kawamura, and A. Adachi. 1995. Function of human immunodeficiency virus type 1 Vpu protein in various cell types. *J. Gen. Virol.* **76**:2717–2722.

Schubert, U., T. Schneider, P. Henklein, K. Hoffmann, E. Berthold, H. Hauser, G. Pauli, and T. Porstmann. 1992. Human immunodeficiency virus type 1 encoded Vpu protein is phosphorylated by casein kinase II. *Eur. J. Biochem.* **204**:875–883.

Schubert, U., and K. Strebel. 1994. Differential activities of the human immunodeficiency virus type 1-encoded Vpu protein are regulated by phosphorylation and occur in different cellular compartments. *J. Virol.* **68**:2260–2271.

Schubert, U., K. A. Clouse, and K. Strebel. 1995. Augmentation of virus secretion by the human immunodeficiency virus type 1 Vpu protein is cell type independent and occurs in cultured human primary macrophages and lymphocytes. *J. Virol.* **69**:7699–7711.

Schubert, U., S. Bour, A. V. Ferrer-Montiel, M. Montal, F. Maldarelli, and K. Strebel. 1996a. The two biological activities of human immunodeficiency virus type 1 Vpu protein involve two separable structural domains. *J. Virol.* **70**:809–819.

Schubert, U., A. V. Ferrer-Montiel, M. Oblatt-Montal, P. Henklein, K. Strebel, and M. Montal. 1996b. Identification of an ion channel activity of the Vpu transmembrane domain and its involvement in the regulation of virus release from HIV-1-infected cells. *FEBS Lett.*, in press:

Schwartz, S., B. K. Felber, E. M. Fenyö, and G. N. Pavlakis. 1990. Env and vpu proteins of human immunodeficiency virus type-1 are produced from multiple bicistronic mRNAs. *J. Virol.* **64**:5448–5456.

Strebel, K., T. Klimkait, and M. A. Martin. 1988. A novel gene of HIV-1, vpu, and its 16-kilodalton product. *Science* **241**:1221–1223.

Strebel, K., T. Klimkait, F. Maldarelli, and M. A. Martin. 1989. Molecular and biochemical analyses of human immunodeficiency virus type 1 vpu protein. *J. Virol.* **63**:3784–3791.

Terwilliger, E. F., E. A. Cohen, Y. Lu, J. G. Sodroski, and W. A. Haseltine. 1989. Functional role of human immunodeficiency virus type 1 vpu. *Proc. Natl. Acad. Sci. USA* **86**:5163–5167.

Theodore, T. S., G. Englund, A. Buckler-Whitte, C. E. Buckler, M. A. Martin, and K. W. C. Peden. 1996. Construction and characterization of a stable full-length macrophage-tropic HIV type 1 molecular clone that directs the production of high titers of progeny virions. *AIDS Res Hum Retrovir* **12**:191–194.

Trono, D., 1995. HIV accessory proteins: Leading roles for the supporting cast. *Cell* **82**:189–192.

Vincent, M. J., N. U. Raja, and M. A. Jabbar. 1993. Human immunodeficiency virus type 1 Vpu protein induces degradation of chimeric envelope glycoproteins bearing the cytoplasmic and anchor domains of CD4: Role of the cytoplasmic domain in Vpu-induced degradation in the endoplasmic reticulum. *J. Virol.* **67**:5538–5549.

Vincent, M. J., and M. A. Jabbar. 1995. The human immunodeficiency virus type 1 Vpu protein: a potential regulator of proteolysis and protein transport in the mammalian secretory pathway. *Virology* **213**:639–649.

Westervelt, P., T. Henkel, D. B. Trowbridge, J. Orenstein, J. Heuser, H. E. Gendelman, and L. Ratner. 1992. Dual regulation of silent and productive infection in monocytes by distinct human immunodeficiency virus type 1 determinants. *J. Virol.* **66**:3925–3931.

Willey, R. L., F. Maldarelli, M. A. Martin, and K. Strebel. 1992a. Human immunodeficiency virus type 1 Vpu protein regulates the formation of intracellular gp160-CD4 complexes. *J. Virol.* **66**:226–234.

Willey, R. L., F. Maldarelli, M. A. Martin, and K. Strebel. 1992b. Human immunodeficiency virus type 1 vpu protein induces rapid degradation of CD4. *J. Virol.* 66 :7193–7200.

Willey, R. L., A. Buckler-White, and K. Strebel. 1994. Sequences present in the cytoplasmic domain of CD4 are necessary and sufficient to confer sensitivity to the human immunodeficiency virus type 1 Vpu protein. *J. Virol.* **68**:1207–1212.

Wray, V., T. Federau, P. Henklein, S. Klabunde, O. Kunert, D. Schomburg, and U. Schubert. 1995. Solution structure of the hydrophilic region of HIV-1 encoded virus protein U (Vpu) by CD and 1H NMR spectroscopy. *Int. J. Peptide Protein Res.* **45**:35–43.

Yao, X. J. Y., H. Göttlinger, W. A. Haseltine, and E. A. Cohen. 1992. Envelope glycoprotein and CD4 independence of Vpu-facilitated human immunodeficiency virus type 1 capsid export. *J. Virol.* **66**:5119–5126.

Yao, X. J., S. Garzon, F. Boisvert, W. A. Haseltine, and E. A. Cohen. 1993. The effect of vpu on HIV-1-induced syncytia formation. *J. Acq. Immune. Defic. Syndr.* **6**:135–141.

Yao, X. J., J. Friborg, F. Checroune, S. Gratton, F. Boisvert, R. P. Sekaly, and E. A. Cohen. 1995. Degradation of CD4 induced by human immunodeficiency virus type 1 Vpu protein: a predicted alpha-helix structure in the proximal cytoplasmic region of CD4 contributes to Vpu sensitivity. *Virology* **209**:615–623.

Genetic Subtypes of HIV-1

Thomas Leitner

Theoretical Biology and Biophysics, Group T-10, MS K710,
Los Alamos National Laboratory, Los Alamos NM 87545

Current Subtypes

HIV-1's are currently divided into two genetic groups, based on phylogenetic reconstruction using DNA sequences. The majority of these sequences fall into the M (major) group, while a smaller, but growing, number of sequences are classified as O (outlier). The M group has been further subdivided into several subtypes formed by more or less distinct clades in the M group phylogeny. These clades have been given subtype classification names from A to J (Figure 1).

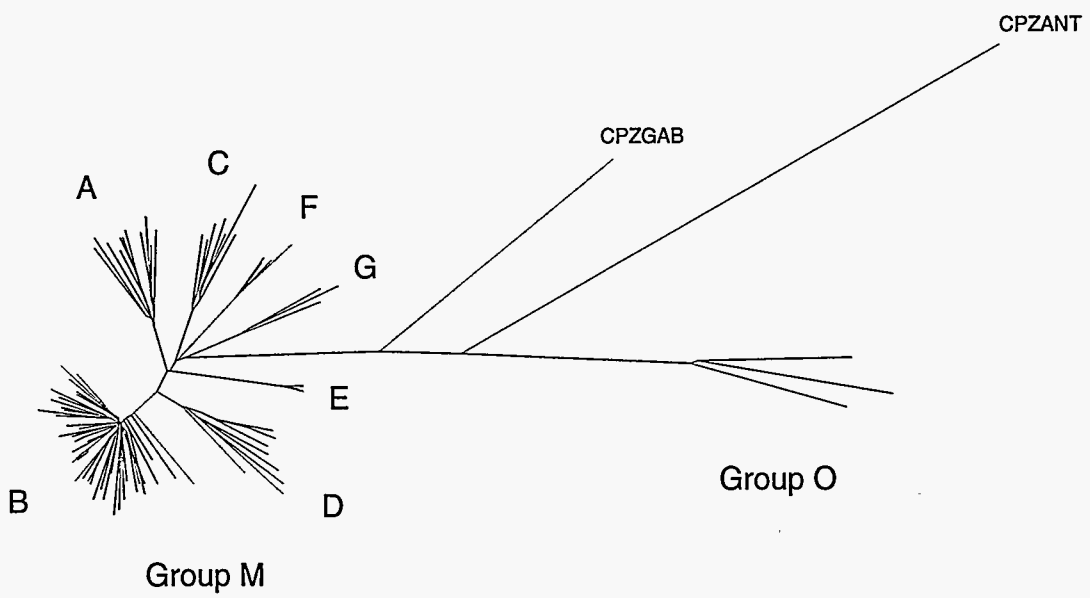


Figure 1. Unrooted neighbor-joining tree showing the HIV-1 group M and O relation with the chimpanzee SIV. The group M subtypes A–G are also indicated at the respective cluster, while for subtypes H and J full length sequences do not yet exist and therefore they are not included in this tree. The tree was based on all nucleotide sequences corresponding to the full length gp120 coding region of *env* in the data base, using a distance matrix calculated with the F84 ($\kappa = 2.0$) substitution model [19].

In phylogenetic trees the chimpanzee sequences usually have a branch point located between Group M and O (Figure 1), suggesting distinct origins of the Group M and O viruses. The branching order of the two chimpanzee sequences (CPZGAB and CPZANT) may change when different sequence fragments are investigated. The explanation for this inconsistency may be recombination between chimpanzee sequences.

Definition of a Subtype

The bases for what should be called a genetic subtype were originally proposed by Myers in *Human Retroviruses and AIDS 1994* [47].

i) subtypes are approximately equidistant from one another in *env* (a "star phylogeny");

ii) the *env* phylogenetic tree is for the most part congruent with the *gag* phylogenetic tree;

iii) two or more samples are required to define a sequence subtype.

These are both logical and reasonable practical criteria for the subdivision of sequences based on phylogenetic trees. However, many sequences were found to be unclassified because they clustered differently using different phylogenetic reconstruction methods or in different reference data sets. Furthermore, as discussed below, some sequences were either too short or were actually just one representative of a potential subtype. With recent knowledge and methods of detecting intersubtype recombinants, many of these unclassified sequences have been explained. In fact, many of these problems are typical for chimeric sequences. Salminen et al., therefore, in the 1995 issue of the *Human Retroviruses and AIDS* [55] suggested that the criteria for new subtypes should include:

i) the appearance of at least two epidemiologically unrelated isolates that cluster together and are separated from established genotypes;

ii) the availability of at least 1.5 kilobases of contiguous sequence from each;

iii) the absence of any subsegment that can join established genotypes.

The sum of all these criteria form a reasonable sound basis for establishing a new subtype. However, when attempting to classify a sequence as an established subtype, less strict criteria may be used, as long as this is made clear. For instance, if only the *env* V3 region (typically 300 bp) was used to classify a virus one might propose that the virus is *env* V3 subtype X (where X = A, B, C, etc.). It would be better, of course, if more of the above criteria were fulfilled. Should the sequence fall outside any of the established clusters, it is desired that further sequence be generated and that additional analyses be performed, as described below.

The M Group

Table 1 lists typical genotypes from each of the group M HIV-1 subtypes. The sequences in this table were chosen accordingly to these criteria: (1) The sequence from the sample should exist in *env*, *gag*, and *pol*; (2) The sequence should be a full length gene sequence; (3) The sequences should representthe variation within each subtype; (4) The sequences should not show any signs of recombination in any of the three genes *env*, *gag*, or *pol*; and (5) A maximum of four sequences should represent each subtype. These criteria were chosen in order to facilitate the subtyping effort. For instance, criterion 1 makes it easier to compare results from different genomic fragments, and criterion 2 makes it possible to use the same references even though studies are based on different subfragments. Criteria 3 and 4 allow a researcher to detect odd sequences for further analysis (see below), and finally, criterion 5 lessens the computational effort required. However, these criteria were not possible to follow in every case: in the more uncommon subtypes (G, H, and J) complete genes have not always been sequenced. Subtype I has not been included in the list, because it is only determined in the *env* V3 region [30], and has also preliminary been proposed to be a recombinant [53]. In *pol*, less sequencing has been done; therefore, the criteria were followed less strictly for this gene. Subtype E is also an exception from criterion 1 [23].

Table 1 also gives a brief description of the principal geographical locales of each subtype. All subtypes have been found in Central Africa, while only one or a few subtypes dominate in other parts of the world. However, in some European countries several subtypes have been reported [4, 21, 30, 54], and in one case nearly all subtypes have been found [1].

Subtype A is the most genetically diverged subtype, and in many tree analyses the more rare subtypes G and H are close to subtype A. Furthermore, all subtype E and at least three subtype G viruses have been found to cluster in subtype A in the 3' half of their gp41 coding region, suggesting that they are recombinants [23]. The Thai envelope subtype E sequences all appear to be mosaic, with

Table 1 Proposed Reference Sequences of HIV-1 Genetic Subtypes

Subtype	*env*	*gag*	*pol*	Main geographical area
A	U455	U455	U455	Central Africa
	IBNG	IBNG	IBNG	
	DJ258	DJ258		
	SF1703	VI32		
B	LAI	LAI	LAI	Europe
	JRFL	JRFL	JRFL	North and South America
	OYI	OYI	OYI	Australia
	RF	RF	RF	Asia
C	UG268	UG268		East and South Africa
	ZAM18	ZAM18		
	ZAM20	ZAM20		
	DJ256	DJ259		
D	NDK	NDK	NDK	Central Africa
	JY1	VI203	Z2Z6[1]	
	UG274	UG274	ELI[1]	
	SE365	UG270		
E	TN235			East Asia
	TN239			
	TN242			
	CM240X			
F	BZ163A	BZ162		South America
	BZ126A	VI69		East Europe
	93BR029.2	VI174		
	93BR020.17			
G	LBV217[1]	LBV217[1]		Central Africa
	92UG975.10	VI191		
	92RU131.9	SE6165[2]	SE6165[2]	
H	VI557[2]	VI557		Central Africa
	CA13[2]	VI525		
J	SE702[2]	SE7022[2]		Central Africa
	SE7887[2]	SE7887[2]		(Europe)

[1] These sequences have been found to be partly recombinant in *env* or *gag*.

[2] Full length gene sequences do not yet exist.

the *gag* gene from subtype A and the *env* from subtype E. The parental subtype E virus has yet not been found, but recently an extensive analysis of one such mosaic isolate showed multiple breakpoints between the two parental genotypes [9]. Subtype A is mostly prevalent in Central Africa, but has also been found in other parts of the world, including Europe [1, 4], Russia [41], East Asia [65], and America [7, 43]. It is possible that this subtype represents a contemporary pool of the original cross-species transmission(s) from monkey to human, from which all the other subtypes have been drawn [34].

Subtype B is by far the most studied subtype, mainly because viruses from this subtype represent the vast majority circulating in the western world, but also because many original laboratory isolates are of this type. However, subtype B has also established successful epidemics in Asia and South America, and is the dominant type in Australia. Most likely has this subtype spread to these parts of the world from contacts with America and Europe. In Thailand a separate type of subtype B has been referred to as B' [26], and more recently B_B , which has been found to further spread into China, Malaysia, and Japan [10]. Subtype B and subtype E transmissions in Thailand have been compared, and some authors have proposed that in this community subtype E was found to be more transmissible [26, 32, 62]. This observation alarmed some European countries, where it was shown that subtype E had recently entered [2]. Subtype E has also entered the American continent, including the USA [5, 7], and subtype E variants have been found in Central Africa [44, 48].

Subtype C seems to have its major center in East and South Africa, but has been reported to occur in many other parts of the world, including Europe, Russia, China, India, and Brazil. As an effect of communication, geographical patterns are becoming more and more blurred for essentially all subtypes. This in turn will increase the possibility of superinfections of different subtypes, and thereby the virus will have greater opportunities to explore new genetic variants by recombination [46].

The O Group

Virus and sequences from the O group have been investigated less thoroughly, mainly because this type of virus is less common. Currently, 3 full length *env* gp120 sequences have been published and are available in the database, but several groups are investigating these viruses. Most of these viruses originate from individuals who have been infected in West Africa, or from encounters with someone infected in West Africa [27]. Analysis of data suggest that the O group sequences do not form subtype clusters in the same way as the M group sequences [39].

Methods of Investigation

Strategy. In any attempt to do a sequence study a good strategy is required. Depending on what the goal is, different approaches will apply. In many cases several areas of interest may have been pursued simultaneously, and in other cases the sequences may have been determined for a goal other than the present one. For instance, samples may have been collected from a restricted area or disease group, or may have gone through different in vitro manipulations; and sequences may have been derived from either proviral DNA or genomic RNA. All these factors should be considered before going further, since they may bias the conclusions.

Sequencing. After an investigation strategy has been decided and samples have been collected, a study often starts with determining the primary DNA sequence. As mentioned above, the sequence template may be either proviral DNA or reverse transcribed RNA. It is important to remember that any HIV sample contains a population of related but genetically diverse variants, or quasispecies [42] (see also [15, 16, 25]). Any sequencing strategy must take this fact into account, i.e. it must be certain that a representative number of viral molecules are investigated. If this is not done, the chance of making erroneous conclusions arises: for instance, one could detect a "homogeneous" population, or a special sequence, derived from a single odd molecule. Thus, the number of molecules that goes into the sequencing effort should be determined, for example by a limiting dilution procedure [60].

The easiest way to determine a virus sequence is by direct PCR sequencing. This procedure suffers the least from methodological artifacts by introducing the least amount of undesired selection steps. Because a population is sequenced, some nucleotide positions may be populated by more than

one state (a multistate position); for instance, a position can harbor both an A and a G. There are different ways to handle this information. Some authors simply choose the dominant form at each such site (e.g., [49]), while others choose to report both nucleotides to a certain threshold [37]. It has also been shown that the different nucleotides at such positions can be accurately quantified [37]. The main drawback of the direct population sequencing strategy is that the linkage between nucleotides in multistate positions will remain unknown in most cases.

The other way to determine a population is by cloning prior to sequencing. There are several ways to do this, but the two most common ways are vector cloning or limited dilution of PCR fragments. Vector cloning involves many steps in which the population structure may be skewed, including transformation, growth, and clone picking. The limited dilution procedure [60] avoids these steps, but instead requires many PCR reactions. Both procedures also suffer from the fact that PCR-induced mutations may be amplified and accidentally analyzed as true information. Another seldom addressed problem with clones is that they generate a much larger dataset. This may be cumbersome in the stage of phylogenetic inference, because the number of possible tree topologies quickly reaches astronomical numbers, thereby making more advanced methods out of the question. However, some studies require exact knowledge of individual molecules, and a cloning procedure becomes the method of choice.

In addition to sequencing, a number of other procedures exist to determine genotypes. These include restriction enzyme analysis, SSCP, and different probe protocols such as the heteroduplex mobility assay (HMA) [12]. The advantage of such methods is that they can screen large sample sets in short time; efforts have been made to produce reference sets for standardization of the HMA method [6]. The resolution of this method is, however, lower than that of a sequencing study. Methods for rapid screening of subtypes based on V3 peptide serotyping have also been developed [58].

An important factor to consider is the amount of information fed to the subsequent analysis, usually a tree reconstruction. Although the discussion of phylogenetic information is difficult, in principle, two components decide whether it will be a good reconstruction or not. The first is sequence length. It is evident that the ideal procedure would be to use the complete HIV genome. However, to sequence such a fragment is difficult, and therefore the sequences of subfragments are usually determined. The second component is information density. The env gene is known to display a much higher degree of variability than the *pol* gene. Thus, a 300 bp fragment covering a hypervariable region in *env* contains far more information than the same length of sequence in pol. Several other factors, such as convergent changes as a result of influence of the immune system and/or drug treatment will affect the quality of the information, but normally, if the length of the sequence is reasonably long, this will only yield a slight background noise. However, it may be wise to choose a region that does not contain extensive length variations, such as V1-V2 of *env*, because much information will be lost in alignment columns where these occur as a result of gapstripping procedures [33] that many phylogenetic inference methods include. Leitner et al. recently showed that a 285 bp fragment covering the *env* V3 region was sufficient to reconstruct a known intrasubtype transmission chain covering 13 years of viral evolution [35]. For classifying subtypes, less would be required because of the larger sequence differences. Figure 2 shows the genetic distance correlation of common subfragments to their complete gene sequences. As is evident from the figure, shorter sequences are essentially giving the same answer as longer ones. This was true for both intra and inter subtype relations. Interestingly, some of the subfragments carry a higher information density than the complete genes: the V3 fragment of *env* and the p17 fragment of *gag* both display longer genetic distances than the complete genes (Figures 2B and 2C). Note, however, that the shorter a fragment gets, the more it will be affected by mutational noise and the lower the significance will be. Yet another problem with too-short sequences is that there may not be enough information there to estimate some parameters for more advanced distance estimates.

Phylogenetic inference. When a DNA sequence has been determined for a sample, the next step is to align it to the set of reference sequences describing established subtypes. Alignments of sequences covering the complete *env*, *gag*, and *pol* genes as well as subfragments gp120 and V3 of *env*, p17 (MA) of *gag*, and p51 (RT) of *pol* of the suggested reference sequences (Table 1) are available at the HIV database (http://hiv-web.lanl.gov). For the complete gene sequences, sets containing the subtypes shown in Figure 3 are included, while for the subfragments also more uncommon subtypes are represented when such sequences were available (according to Table 1).

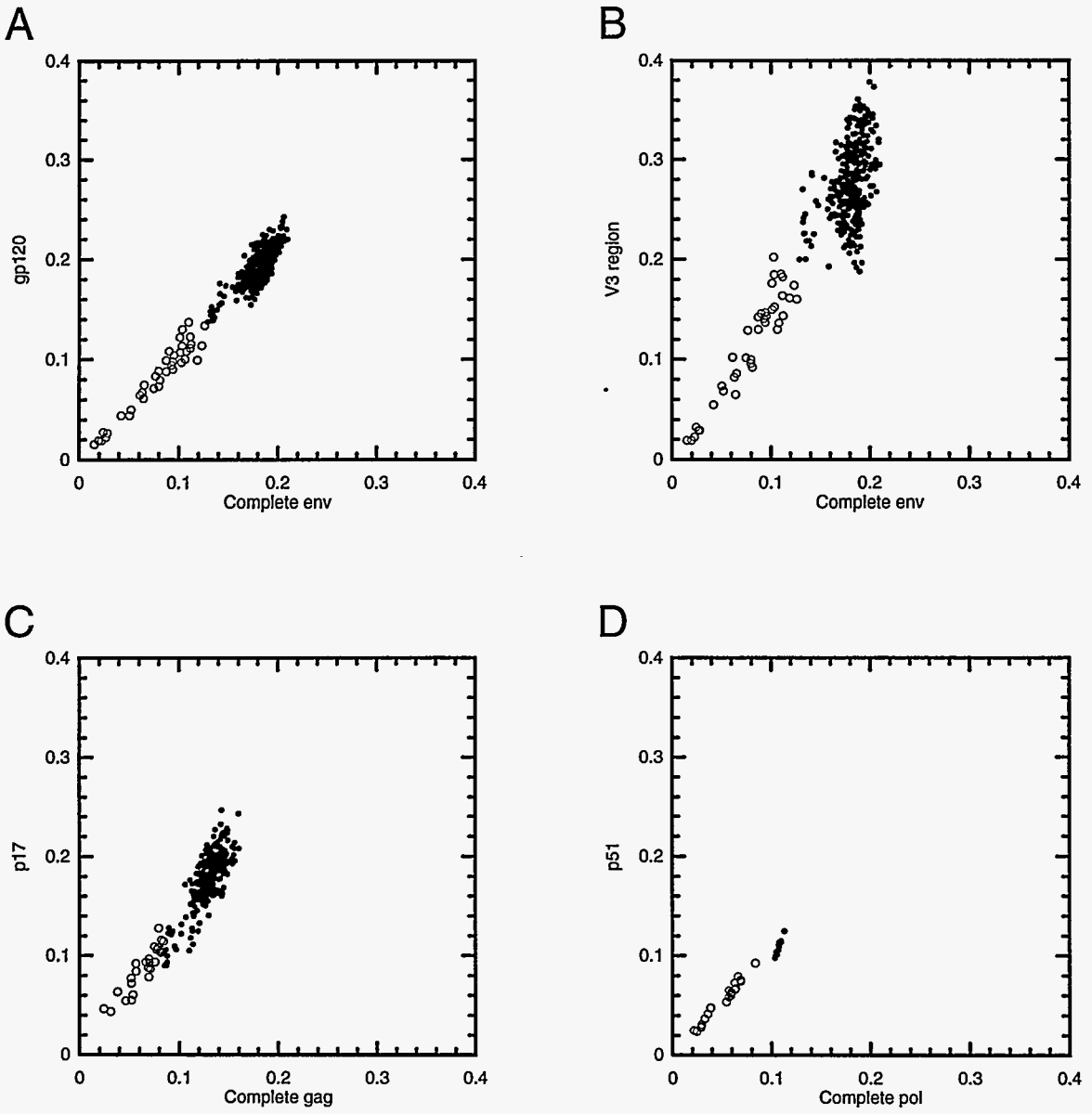


Figure 2. Comparison of distance estimates of nucleotide sequences from complete gene fragments to common gene subfragments. Open circles indicate within-subtype distances, and solid circles between-subtype distances. (A) The gp120 coding region (1251 bp) compared to the complete *env* gene (2301 bp); (B) The V3 region of gp120 (324 bp) compared to the complete *env* gene; (C) The p17 fragment of *gag* (381 bp including 24 bp into p24) compared to the complete *gag* gene (1380 bp); (D) The p51 fragment of *pol* (1320 bp) compared to the complete pol gene (3000 bp). All distance estimates were calculated using the F84 ($\kappa = 2.0$) nucleotide substitution model [19]. Note that sequence lengths are given for globally gapstripped datasets.

Figure 3 shows how maximum-likelihood, maximum-parsimony, and minimum-evolution based phylogenetic methods perform in reconstructing the established subtypes by use of *env*, *gag*, or *pol* sequences. All these methods and fragments can reconstruct the established subtypes. These trees were all calculated using the full length gene sequences, but essentially the same result would have been

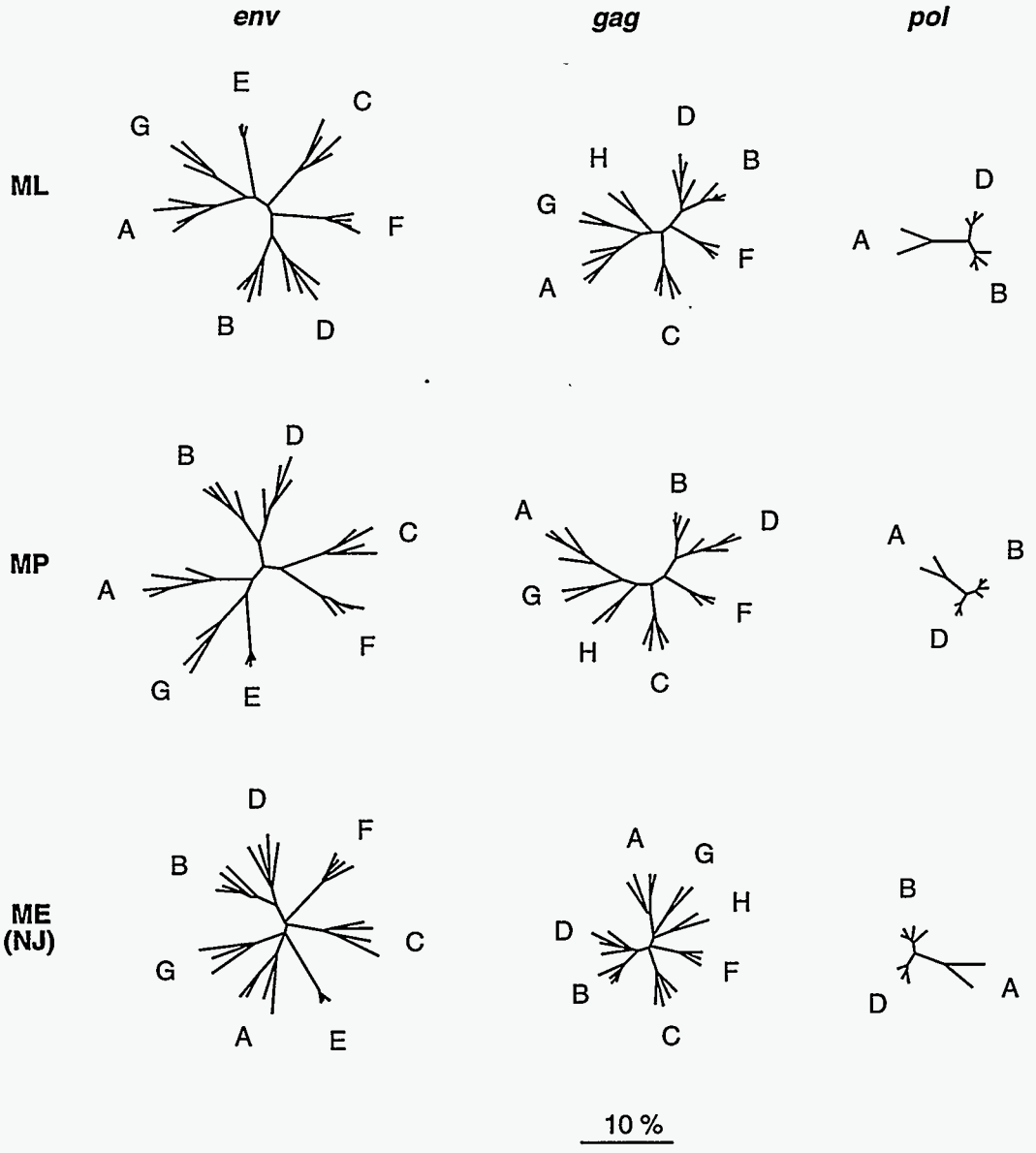


Figure 3. Unrooted trees calculated according to the maximum likelihood (ML) criteria, the maximum parsimony (MP) criteria, and the minimum evolution (ME) criteria. The ML search was conducted using the PHYLIP package [19] with the F84 substitution model, the MP search was conducted with a heuristic search using the PAUP program [63] under uniform weighting, and the ME search was conducted with the neighbor joining approximation using the PHYLIP package [19] with the F84 substitution model. Each method was applied to the complete *env*, *gag*, and *pol* gene alignments, gapstripped to 2301, 1380, and 3000 nucleotide sites, respectively. Branch lengths in ML and ME trees were adjusted to the indicated scale bar, and MP trees were drawn arbitrarily to reflect the information content in a comparable scale.

derived using the suggested subfragments (Figure 2). Thus, for subtyping purposes methods based on these criteria (maximum-likelihood, maximum-parsimony, and minimum-evolution) will give an acceptable answer. In other situations the method may influence the result. However, it has recently been shown in detailed analyses of transmission cases that the amount of information is more important

than the choice of phylogenetic inference method [35]. The opposite is true for calculating genetic distances [38]. Although this is not a primary goal when attempting to determine the subtype of a sequence, it may provide useful information regarding relatedness. When a p-distance (proportion of differences) reaches above 0.1, the choice of substitution model is already making a major impact on the distance estimates. From here, the effect of superimposed events is then rapidly increasing by increased distance. By using an oversimplified model of nucleotide substitution, longer distances will be less and less accurate. With increased distance, saturation effects also start to appear, especially on synonymous positions. If the sequence is long enough, one may consider using only nonsynonymous positions, only second codon positions, or translated amino acid sequences. Again, however, for simple subtyping purposes the substitution model is unlikely to influence the results.

Figure 3 also indicates the information density of the three investigated genes. As is evident from all trees, independent of phylogenetic method, the branch lengths of the tree are not reflected by the sequence length, but rather by the density of nucleotide variation of each gene. In other words, the faster a gene evolves the longer the branch lengths will be, as compared to the more slowly evolving gene over the same time interval. The speed at which the genes are evolving is dictated by several factors, most notably mutation and selection.

Rooting or using a specific sequence as an outgroup in a tree analysis, makes no difference when determining the subtype of a sequence. Depending on which type of tree plot one chooses to present the results, however, the choice of outgroup may influence the aesthetics of the tree. A root should only be assessed when its location in the phylogeny is known, while an outgroup can be chosen arbitrarily [64].

When a sequence falls outside an established subtype, more analyses are needed. The first thing to do would be to check if the used sequence fragment includes any sign of recombination. This can be done by several methods, including signature analysis by VESPA [17, 28] or RIP [59], tree analysis by the bootscanning method [56], or even by simply eyeballing an alignment against representative subtype sequences or subtype consensus sequences. A more extensive tree analysis should also be done, including all available sequences from the two subtypes surrounding the odd sequence as well as earlier reported unclassified sequences. The next step would be to sequence another part of the HIV genome, to see if the classification there indicates an odd location as well. Congruence of trees from two separate coding regions is a strong indication of a new genotype. Also, one should check that the sequence analyzed is not odd because of hypermutation events [66]. If the sequence falls outside established subtypes in two regions, and no signs of recombination are seen, it would be desirable to sequence the complete genome to allow for extensive recombination analysis and/or to establish a new subtype (or potential subtype, if only one sample was found).

Bootstrapping is a resampling technique used to test the robustness of an inferred tree topology [13, 18]. The method resamples data points from the original dataset, and for each resampled dataset a tree is inferred. Finally a consensus tree is constructed and each cluster in this tree is assigned a "bootstrap value" corresponding to the number of times it occurred in the resampled trees. The interpretation of such values varies among authors, since the values provide unbiased but highly imprecise estimates of repeatability. Some authors have claimed that the probability that the result represents the true phylogeny is a biased, but usually conservative, estimate of accuracy [24]. Recently, however, other authors showed that the method is not biased, but that it can be corrected to better agree with standard ideas of confidence levels and hypothesis testing [14]. The degree of "bias" varies from node to node, as well as from study to study, so bootstrap values are not directly comparable to each other. Additionally, bootstrap values for a given cluster in any tree will be highly affected by the control sequences chosen. However, many subtype studies include this analysis, and at best it gives a measure of how stable the inference is. In general, all the established subtypes should present high bootstrap values, but congruence of trees from two coding regions are preferable to bootstrap analysis over just one region.

Implications and Epidemiology

Genetic subtypes were originally defined on *env* and *gag* sequences. The main goal was to identify the major genetic lineages [40, 45]. Since then many studies have included or even concentrated on efforts of subtype classifications. The presence of distant genotypes also made it possible to investigate and realize that recombination events (and thereby, superinfections) are common, or at least not uncommon, in HIV evolution [36, 50, 52]. However, the key question that has been raised among researchers in this field is: What does a genetic subtype mean?

There are three main reasons for why genetic subtypes have been considered to be important to investigate, where the first two are based on the belief that the viral genetic factor is important in virus to host interactions. First, in the context of vaccine development it was realized that the different subtypes express different envelope proteins, and that this likely would effect vaccine treatments. For instance, the tip of the principal neutralization domain of the V3 loop is GPGR in subtype B viruses, while in most other subtypes the consensus is GPGQ. Together with the knowledge that genetic variants can escape the immune system, information on genetic subtypes becomes important [8, 26]. Another similar problem, resulting from genetic variation among subtypes is that some methods can fail in detecting virus [3].

Second, several studies have tried to correlate genetic factors to biological properties. For instance, specific amino acids in the *env* V3 loop have been found to be markers for the viral phenotype [11, 20]. However, such a correlation have been found to be subtype independent, or in other words, different phenotypes exist in all genetic subtypes [36, 51, 68]. Kostrikis et al. recently showed that no direct correlation between neutralization serotypes and genetic subtypes could be established [31]. It has been suggested that the Thai subtype E may be more transmissible than subtype B in Thailand [32, 62], but it is difficult to discriminate between the viral genetic factor and the epidemiological situation in such a comparison. Recently, viral phenotype was found to be dominant over genotype also in the context of chemokine receptors as being the long-sought second receptor of HIV-1 [67]. Genotypic studies have often concentrated to regions of special interest, and it is therefore not clear if genetic factors elsewhere in the HIV genome may play an important biological role. However, mutational trends have been observed within the V3 loop among different subtypes, which may reflect the acquisition of specific biological properties during the evolution of HIV-1 [29].

Finally, more informative are the subtypes for the understanding of the HIV-1 epidemic. It was early realized that subtype B represented the vast majority of virus circulating in North America and Europe, while it was almost absent in Africa. Instead, all other subtypes were found in Africa (more recently, subtype B has been reported several times from different African countries). Together with the fact that many species of African monkeys carry similar viruses, it was concluded that the origin of HIV was somewhere in Africa. The question of when the cross-species transmission from monkeys to humans occurred has been investigated by several groups, and the answers have been astonishingly different: ranging from 40 to millions of years ago [22, 57, 61]. The main reason for the different interpretations is that the true evolutionary pattern of the virus is not known, creating problems of interpreting the shape of the tree and correlating time to branch lengths. However, the current subtypes do tell us that very different viral genotypes can rapidly establish successful epidemics in naive human populations. This suggests that each subtype is well-adapted to its environment, but also that any existing variant could have established a successful clade. In this context, it is interesting to look a bit closer to the actual shape of the subtype tree (Figures 1 and 3). We can see that the star-like shape actually has structure: subtype B and D are always closer to each other; subtype A is usually more divergent than other subtypes and normally close to G and H; and that more recent subtypes form more distinct clusters away from clusters that mainly are derived from Central African sequences. It seems like there are two components in the tree: an early phase in which the virus spread more slowly between host groups in Africa, forming its network of variants; and a recent phase, which we are still in, in which the virus reaches a naive human population and uses its standing network to spread rapidly. Perhaps can this teach us about epidemic spreads, and how to prevent them in the future.

References

[1] Alaeus, A., T. Leitner, K. Lidman, and J. Albert. 1996. Most genetic subtypes of HIV-1 have entered Sweden. *AIDS*:in press.

[2] Anonymous. 1996. Also HIV subtype E is not transmitted by kissing. *Z. Arztl. Fortbild. (Jena).* **90**:49 (in German).

[3] Arnold, C., K. L. Barlow, S. Kaye, C. Loveday, P. Balfe, and J. P. Clewley. 1995. HIV type 1 sequence subtype G transmission from mother to infant: failure of variant sequence species to amplify in the Roche Amplicor Test. *AIDS Res. Hum. Retrovirus.* **11**:999–1001.

[4] Arnold, C., K. L. Barlow, J. V. Parry, and J. P. Clewley. 1995. At least five HIV-1 subtypes (A, B, C, D, A/E) occur in England. *AIDS Res. Hum. Retrovirus.* **11**:427–429.

[5] Artenstein, A. W., C. A. Ohl, T. C. VanCott, P. A. Hegerich, and J. R. Mascola. 1996. Transmission of HIV-1 subtype E in the united states. *JAMA.* **276**:99–100.

[6] Bachmann, M. H., E. L. Delwart, E. G. Shpaer, P. Lingenfelter, R. Singal, and J. I. Mullins. 1994. Rapid genetic characterization of HIV type 1 strains from four World Health Organization-sponsored vaccine evaluation sites using a heteroduplex mobility assay. WHO Network for HIV Isolation and Characterization. *AIDS Res. Hum. Retrovirus.* **10**:1345–1353.

[7] Brodine, S. K., J. R. Mascola, P. J. Weiss, S. I. Ito, K. R. Porter, A. W. Artenstein, F. C. Garland, F. E. McCutchan, and D. S. Burke. 1995. Detection of diverse HIV-1 genetic subtypes in the USA. *Lancet.* **346**:1198–1199.

[8] Buonaguro, L., E. Del Guadio, M. Monaco, D. Greco, P. Corti, E. Beth-Giraldo, F. M. Buonaguro, and G. Giraldo. 1995. Heteroduplex mobility assay and phylogenetic analysis of V3 region sequences of human immunodeficiency virus type 1 isolates from Gulu, northern Uganda. The Italian-Ugandan Cooperation AIDS Program. *J. Virol.* **69**:7971–7981.

[9] Carr, J. K., M. O. Salminen, C. Koch, D. Gotte, A. W. Artenstein, P. A. Hegerich, D. St. Louis, D. S. Burke, and F. E. McCutchan. 1996. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J. Virol.* **70**:5935–5943.

[10] Cassol, S., B. G. Weniger, P. G. Babu, M. O. Salminen, X. Zheng, M. T. Htoon, A. Delaney, M. O'Shaughnessy, and C.-Y. Ou. 1996. Detection of HIV type 1 env subtypes A, B, C, and E in Asia using dried blood spots: a new surveillance tool for molecular epidemiology. *AIDS Res. Hum. Retrovirus.* **12**:1435–1441.

[11] De Jong, J. J., A. De Ronde, W. Keulen, M. Tersmette, and J. Goudsmit. 1992. Minimal requirements for the human immunodeficiency virus type 1 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. *J. Virol.* **66**:6777–6780.

[12] Delwart, E. L., E. G. Shpaer, J. Louwagie, F. E. McCutchan, M. Grez, H. Rübsamen-Waigmann, and J. I. Mullins. 1993. Genetic relationships determined by a DNA heteroduplex mobility assay: analysis of HIV-1 env genes. *Science.* **262**:1257–1261.

[13] Efron, B. 1982. The jackknife, the bootstrap, and other resampling techniques. CBMS-NFS Regional Conference Series in Applied Mathematics, Monograph 38. Society of Industrial and Applied Mathematics, Philadelphia.

[14] Efron, B., E. Halloran, and S. Holmes. 1996. Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. USA.* **93**:13429–13434.

[15] Eigen, M., J. McCaskill, and P. Schuster. 1988. Molecular quasi-species. *J. Phys. Chemis.* **92**:6881–6891.

[16] Eigen, M., and R. Winkler-Oswatitsch. 1990. Statistical geometry in sequence space. *Meth. Enymol.* **183**:505–530.

[17] Farmer, A., and G. Myers. 1995. Detection of HIV hybrid sequences using VESPA. In G. Myers and B. Korber and B. H. Hahn and K.-T. Jeang and J. W. Mellors and F. E. McCutchan and L. E. Henderson and G. N. Pavlakis (ed.), *Human Retroviruses and AIDS 1995: A compilation of nucleic acid and amino acid sequences.* Los Alamos National Laboratory, Los Alamos, NM.

[18] Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution.* **39**:783–791.

[19] Felsenstein, J. 1993. PHYLIP: Phylogeny Inference Package, 3.52c ed. University of Washington, Seattle, WA.

[20] Fouchier, R. A. M., M. Groenink, N. A. Kootstra, M. Tersmette, H. G. Huisman, F. Miedema, and H. Schuitemaker. 1992. Phenotype-associated sequence variation in the third variable region of the human immunodeficiency virus type 1 gp120 molecule. *J. Virol.* **66**:3183–3187.

[21] Fransen, K., A. Buve, J. N. Nkengasong, M. Laga, and G. van der Groen. 1996. Longstanding presence in Belgians of multiple non-B HIV-1 subtypes. *Lancet.* **347**:1403.

[22] Fukasawa, M., T. Miura, A. Hasegawa, S. Morikawa, H. Tsujimoto, K. Miki, T. Kitamura, and M. Hayami. 1988. Sequence of simian immunodeficiency virus from African green monkey, a new member of the HIV/SIV group. *Nature.* **333**:457–461.

[23] Gao, F., S. G. Morrison, D. L. Robertson, C. L. Thornton, S. Craig, G. Karlsson, J. Sodroski, M. Morgado, B. Galvao-Castro, H. von Briesen, S. Beddows, J. Weber, P. M. Sharp, G. M. Shaw, and B. H. Hahn. 1996. Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. The WHO and NIAID Networks for HIV Isolation and Characterization. *J. Virol.* **70**:1651–1657.

[24] Hillis, D. M., and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* **42**:182–192.

[25] Holland, J. J., J. C. De La Torre, and D. A. Steinhauer. 1992. RNA virus populations as quasispecies. *Curr. Top. Microbiol. Immunol.* **176**:1–20.

[26] Kalish, M. L., A. Baldwin, S. Raktham, C. Wasi, C.-C. Luo, G. Schochetman, T. D. Mastro, N. L. Young, S. Vanichseni, H. Rubsamen-Waigmann , H. von Briesen, J. I. Mullins, E. Delwart, B. Herring, J. Esparza, W. L. Heyward, and S. Osmanov. 1995. The evolving molecular epidemiology of HIV-1 envelope subtypes in injecting drug users in Bangkok, Thailand: implications for HIV vaccine trials. *AIDS.* **9**:851–857.

[27] Korber, B., I. Loussert-Ajakai, J. Blouin, and S. Saragosti. 1996. A comparison HIV-1 M and group O functional and immunogenic domains in the gag p24 protein and the C2V3 region of the envelope protein, p. III-39. Published in this compendium.

[28] Korber, B., and G. Myers. 1992. Signature pattern analysis: a method for assessing viral sequence relatedness. *AIDS Res. Hum. Retrovirus.* **8**:1549–1558.

[29] Korber, B. T. M., K. MacInnes, R. F. Smith, and G. Myers. 1994. Mutational trends in V3 loop protein sequences observed in different genetic lineages of human immunodeficiency virus type 1. *J. Virol.* **68**:6730–6744.

[30] Kostrikis, L. G., E. Bagdades, Y. Cao, L. Zhang, D. Dimitriou, and D. D. Ho. 1995. Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: identification of a new subtype designated subtype I. *J. Virol.* **69**:6122–6130.

[31] Kostrikis, L. G., Y. Cao, H. Hgai, J. P. Moore, and D. D. Ho. 1996. Quantitative analysis of serum neutralization of human immunodeficiency virus type 1 from subtypes A, B, C, D, E, F, and I: lack of direct correlation between neutralization serotypes and genetic subtypes and evidence for prevalent serum-dependent infectivity enhancement. *J. Virol.* **70**:445–458.

[32] Kunanusont, C., H. M. Foy, J. K. Kreiss, S. Rerks-Ngarm, P. Phanuphak, S. Raktham, C.-P. Pau, and N. L. Young. 1995. HIV-1 subtypes and male-to-female transmission in Thailand. *Lancet.* **345**:1078–1083.

[33] Learn, G. H. J., B. T. M. Korber, B. Foley, B. H. Hahn, S. M. Wolinsky, and J. I. Mullins. 1996. Maintaining the integrity of human immunodeficiency virus sequence databases. *J. Virol.* **70**:5720–5730.

[34] Leitner, T. 1996. Genetic variation of HIV-1: Molecular epidemiology and viral evolution. Swedish Institute for Infectious Disease Control, Karolinska Institute, Stockholm, Sweden.

[35] Leitner, T., D. Ecanilla, C. Franzén, M. Uhlén, and J. Albert. 1996. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proc. Natl. Acad. Sci. USA.* **93**:10864–10869.

[36] Leitner, T., D. Escanilla, S. Marquina, J. Wahlberg, C. Broström, H. B. Hansson, M. Uhlén, and J. Albert. 1995. Biological and molecular characterization of subtype D, G and A/D recombinant HIV-1 transmissions in Sweden. *Virology.* **209**:136–146.

[37] Leitner, T., E. Halapi, G. Scarlatti, P. Rossi, J. Albert, E. M. Fenyö, and M. Uhlén. 1993. Analysis of heterogeneous viral populations by direct DNA sequencing. *BioTechniques.* **15**:120–126.

[38] Leitner, T., S. Kumar, and J. Albert. 1996. Tempo and mode of the nucleotide substitution in gag and env gene fragments in human immunodeficiency virus type 1 populations with a known transmission history. Submitted.

[39] Loussert-Ajakai, I., M.-L. Chaix, B. Korber, F. Letourneur, E. Gomas, T.-D. Ly, F. Brun-Vezinet, F. Simon, and S. Saragosti. 1995. The variability of HIV type 1 group O strains isolated from Cameroonian patients living in France. *J. Virol.* **69**:5640–5649.

[40] Louwagie, J., F. E. McCutchan, M. Peeters, T. P. Brennan, B. E. Sanders, G. A. Eddy, G. van der Groen, K. Fransen, G.-M. Gershy-Damet, R. Deleys, and D. S. Burke. 1993. Phylogenetic analysis of gag genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. *AIDS.* **7**:769–780.

[41] Lukashov, V. V., M. T. Cornelissen, J. Goudsmit, M. N. Papuashvilli, P. G. Rytik, R. M. Khaitov, E. V. Karamov, and F. de Wolf. 1995. Simultaneous introduction of distinct HIV-1 subtypes into risk groups in Russia, Byelorussia and Lithuania. *AIDS.* **9**:435–439.

[42] Meyerhans, A., R. Cheynier, J. Albert, M. Seth, S. Kwok, J. Sninsky, L. Morfeldt-Månson, B. Åsjö, and S. Wain-Hobson. 1989. Temporal fluctuations in HIV quasispecies in vivo are not reflected by sequential HIV isolations. *Cell.* **58**:901–910.

[43] Montpetit, M. 1995. HIV-1 subtype A in Canada. *AIDS Res. Hum. Retrovirus.* **11**:1421–1422.

[44] Murphy, E., B. Korber, M. C. Georges-Courbot, B. You, A. Pinter, D. Cook, M. P. Kieny, A. Georges, C. Mathiot, F. Barre-Sinoussi, and M. Girard. 1993. Diversity of V3 region sequences of Human Immunodeficiency viruses type-1 from the Central-African-Republic. *AIDS Res. Hum. Retrovirus.* **9**:997–1006.

[45] Myers, G. 1993. Assimilating HIV sequences. *AIDS Res. Hum. Retrovirus.* **9**:697–702.

[46] Myers, G., B. Korber, B. H. Hahn, K.-T. Jeang, J. W. Mellors, F. E. McCutchan, L. E. Henderson, and G. N. Pavlakis. 1995. *Human retroviruses and AIDS 1995: a compilation and analysis of nucleic acid and amino acid sequences.* Los Alamos National Laboratory, Los Alamos, NM.

[47] Myers, G., B. Korber, S. Wain-Hobson, K.-T. Jeang, L. E. Henderson, and G. N. Pavlakis. 1994. *Human retroviruses and AIDS 1994: a compilation and analysis of nucleic acid and amino acid sequences.* Los Alamos National Laboratory, Los Alamos, NM.

[48] Nkengasong, J. N., W. Janssens, L. Heyndrickx, K. Fransen, P. M. Ndumbe, J. Motte, A. Leonaers, M. Ngolle, J. Ayuk, P. Piot, and G. van der Groen. 1994. Genotypic subtypes of HIV-1 in Cameroon. *AIDS.* **8**:1405–1412.

[49] Ou, C. Y., C. A. Ciesielski, G. Myers, C. I. Bandea, C.-C. Luo, B. T. M. Korber, J. I. Mullins, G. Schochetman, R. L. Berkelman, A. N. Economou, J. J. Witte, L. J. Furman, G. A. Satten, K. A. MacInnes, J. W. Curran, H. W. Jaffe, Laboratory investigation group, and Epidemiologic investigation group. 1992. Molecular epidemiology of HIV transmission in a dental practice. *Science.* **256**:1165–1171.

[50] Robertson, D. L., P. M. Sharp, F. E. McCutchan, and B. H. Hahn. 1995. Recombination in HIV-1. *Nature.* **374**:124–126.

[51] Rübsamen-Waigmann, H., H. von Briesen, H. Holmes, A. Björndal, B. Korber, R. Esser, S. Ranjbar, P. Tomlinson, B. Galvao-Castro, E. Karita, S. Sempala, C. Wasi, S. Osmanov, and E. M. Fenyö. 1994. Standard conditions of virus isolation reveal biological variability of HIV type 1 in different regions of the world. WHO Network for HIV Isolation and Characterization. *AIDS Res. Hum. Retrovirus.* **10**:1401–1408.

[52] Sabino, E. C., E. G. Shpaer, M. G. Morgado, B. T. M. Korber, R. S. Diaz, V. Bongertz, S. Cavalcante, B. Galvão-Castro, J. I. Mullins, and A. Mayer. 1994. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. *J. Virol.* **68**:6340–6346.

[53] Salminen, M. 1996. Personal communication.

[54] Salminen, M., A. Nykänen, H. Brummer-Korvenkontio, M. L. Kantanen, K. Liitsola, and P. Leinikki. 1993. Molecular epidemiology of HIV-1 based on phylogenetic analysis of in vivo gag p7/p9 direct sequences. *Virology.* **195**:185–194.

[55] Salminen, M. O., J. K. Carr, D. S. Burke, and F. E. McCutchan. 1995. Genotyping of HIV-1, p. III30–III34. In G. Myers and B. Korber and B. H. Hahn and K.-T. Jeang and J. W. Mellors and F. E. McCutchan and L. E. Henderson and G. N. Pavlakis (ed.), *Human retroviruses and AIDS 1995: a compilation and analysis of nucleic acid and amino acid sequences.* Los Alamos National Laboratory, Los Alamos, NM.

[56] Salminen, M. O., J. K. Carr, D. S. Burke, and F. E. McCutchan. 1995. Identification of breakpoints in intergenotypic recombinants of HIV-1 by bootscanning. *AIDS Res. Hum. Retrovirus.* **11**:1423–1425.

[57] Sharp, P. M., D. L. Robertson, F. Gao, and B. H. Hahn. 1994. Origins and diversity of human immunodeficiency viruses. *AIDS.* **8** (suppl 1):S27–S42.

[58] Sherefa, K., A. Sonnerborg, J. Steinbergs, and M. Sallberg. 1994. Rapid grouping of HIV-1 infection in subtypes A to E by V3 peptide serotyping and its relation to sequence analysis. *Biochem. Biophys. Res. Commun.* **205**:1658–1664.

[59] Siepel, A. C., A. L. Halpern, C. Macken, and B. Korber. 1995. A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. *AIDS Res. Hum. Retrovirus.* **11**:1413–1416.

[60] Simmonds, P., P. Balfe, J. F. Peutherer, C. A. Ludlam, J. O. Bishop, and A. J. Leigh Brown. 1990. Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. *J. Virol.* **64**:864–872.

[61] Smith, T. F., A. Srinivasan, G. Schochetman, M. Marcus, and G. Myers. 1988. The phylogenetic history of immunodeficiency viruses. *Nature.* **333**:573–575.

[62] Soto-Ramirez, L. E., B. Renjifo, M. F. McLane, R. Marlik, C. O'Hara, R. Sutthent, C. Wasi, P. Vithayasai, V. Vithayasai, C. Apichartpiyakul, P. Auewarakul, V. Peña Cruz, D.-S. Chui, R. Osathanondh, K. Mayer, T.-H. Lee, and M. Essex. 1996. HIV-1 Langerhans' cell tropism associated with heterosexual transmission of HIV. *Science.* **271**:1291–1293.

[63] Swofford, D. L. 1991. PAUP: Phylogenetic Analysis Using Parsimony, 3.1.1 ed. Illinois Natural History Survey, Champaign, Illinois.

[64] Swofford, D. L., and G. J. Olsen. 1990. Phylogeny reconstruction, p. 411–501. In D. M. Hillis and C. Moritz (ed.), *Molecular systematics.* Sinauer Associates, Sunderland, MA.

[65] Tsuchie, H., T. S. Saraswathy, M. Sinniah, B. Vijayamalar, J. K. Maniar, O. T. Monzon, R. T. Santana, F. J. E. Paladin, C. Wasi, P. Thongcharoen, M. M. Hossain, O. Yamada, S. Kageyama, T. Kitamura, and T. Kurimura. 1995. HIV-1 variants in South and South-East Asia. *Int. J. STD AIDS.* **6**:117–120.

[66] Vartanian, J.-P., A. Meyerhans, M. Sala, and S. Wain-Hobson. 1994. G - A hypermutation of the human immunodeficiency virus type 1 genome: evidence for dCTP pool imbalance during reverse transcription. *Proc. Natl. Acad. Sci. USA.* **91**:3092–3096.

[67] Zhang, L., Y. Huang, T. He, Y. Cao, and D. D. Ho. 1996. HIV-1 subtype and second-receptor use. *Nature.* **383**:768.

[68] Zhong, P., M. Peeters, W. Janssens, K. Fransen, L. Heyndrickx, G. Vanham, B. Willems, P. Piot, and G. van der Groen. 1995. Correlation between genetic and biological properties of biologically cloned HIV type 1 viruses representing subtypes A, B, and D. *AIDS Res. Hum. Retrovirus.* **11**:239–248.

A Comparison of HIV-1 Group M and Group O Functional and Immunogenic Domains in the Gag p24 Protein and the C2V3 Region of the Envelope Protein

Bette Korber,[1] Ibtissam Loussert-Ajakai,[2] John Blouin,[1] and Sentob Saragosti[3]

[1] *T10 Los Alamos National Lab, Los Alamos, NM 87545, USA*
[2] *Laboratoire de Virologie, Hôpital Bichat-Claude Bernard, 75877 Paris cedex 18, France*
[3] *I.C.G.M. Cochin Hospital 75014 Paris, France*

The major focus of the sequencing effort of group O viruses to date has been on the p24 Gag and C2V3 gp120 env regions. A total of 42 p24 sequences and 45 C2V3 sequences derived from group O isolates have either been published or provided to the Los Alamos database prior to publication. The primary goals of this review are: i) to provide a brief summary of the accumulating knowledge concerning HIV-1 group O serology and epidemiology; ii) to outline differences and similarities group M and O HIV-1 viruses in the the heavily sequenced 24 Gag and C2V3 gp120 regions, with an emphasis on well-defined functional domains; and iii) to provide an alignment of the region of gp41 that harbors an immunodominant domain important for diagnostics. Only published sequences will be shown in alignments (Table 1), but complete sets of unpublished and published sequences currently in the database were used to generate the consensus sequences for group O sequences shown in the accompanying alignment figures.

Introduction

Group O characterization: After the discovery of the virus ANT70 by a group in Belgium [de Leys (1990)], and the further characterization of this isolate, as well as the MVP5180 and VAU isolates (VAU was isolated from a French woman with no ties to Africa) in 1994 [vanden Haesevelde (1994), Gurtler (1994), Charneau (1994)], it became clear that a new group of viruses, distinct from the HIV-1 Major (M) group, was present in the human population. This group was named the HIV-1 group O for "Outlier" group. The group O, group M, and chimpanzee viruses are genetically quite distant from one another, but these three groups share a distinct lineage relative to other primate lentiviurses (Figure 1). It is notable that four years separated the publications concerning the first and second of these outlier viruses.

Since the isolation of these first highly unusual HIV-1 strains, a second Caucasian patient was found in France [Cohen (1995)], and a quite large number of viruses of this group have been isolated, especially from Cameroonian patients [Loussert-Ajaka (1995)]. The main method for identifying group O viral infections is obtaining a paradoxical pattern of positive and negative serological reactivity using a set of HIV-1 enzyme-linked immunosorbent assay detection kits [Simon (1994)]. A variety of Western blot reactivity patterns can occur with group O serum. Phenotypic differences between group O and M have also been detected. HIV-1 group O isolates exhibited a high level of resistance to non-nucleoside inhibitors ([Descamps (1995)], and Descamps et al., manuscript in preparation), and group M, but not group O viruses, require the incorporation of cyclophilin A for the production of infectious virions [Braaten (1996)].

Sequence analysis of env and gag regions of the first ten group O isolates revealed a peculiar feature of this group relative to group M isolates – limited phylogenetic clustering patterns with only a few sequences grouping into clades in both the p24 and C2V3 regions [Loussert-Ajaka (1995)]. This is in marked contrast to the clear phylogenetic subtypes that emerge in phylogenetic analysis of the group M sequences. The original ten group O isolates are a subset of the 40-plus sequences currently stored in the Los Alamos database, and the observation of limited subtype structure among O group sequences [Loussert-Ajaka (1995)] is consistent with the additional data now available. The nucleotide

Table 1. List of public HIV-1 O group sequences included in this paper. The first two characters preceding the period indicate the country of residence of the individual who was the source of the isolate, followed by the sequence name.

Sequence	Accesion	Reference	Note
O_DE.HAM112		Hampl et al.	The protein sequences were entered by hand at the Los Alamos Database, from the paper by Hampl et al. The source was a 27 year old Cameroonian who had moved to Germany, and was tested for HIV infection because of eczematous skin lesions. This was the first group O infection to be found in Germany.
O_CM.ANT70	L20587	De Leys et al and vanden Haesevelde et al.	This was the first O subtype isolate discovered. The complete viral genome has been sequenced. The virus was derived from a symptomatic Cameroonian, CDC stage III.
O_FR.VAU	X80020	Charneau et al.	This sequence was from an isolate from a French woman who died of AIDS in 1992, who had no known ties to West Africa. DNA was extracted from VAU infected PBMCs, PCR amplified, cloned, and gp160 env was sequenced. The viral isolate was highly cytopathic.
O_CM.MVP5180	L20571	Gurtler et al.	The complete viral genome has been sequenced from the MVP5180 isolate derived from a Cameroonian woman, sampled in 1991; the donor died of AIDS in 1992. The viral isolate MVP5180 was grown in several human T-cell lines and the monocytic U937 line.
O_FR.BCF 01, 02, 03, 06, 06, 08, 11	C2V3 region U24562-U24568, p24 U24706-U24712	Loussert-Ajaka et al.	These seven sequences are from Cameroonian patients who were living in France. PBMC proviral DNA was PCR amplified and 3-6 clones from each patient were sequenced. The consensus of the 3-6 clones was presented. 07, 08 and 11 were CDC stage II; 01 was CDC IV; and 02, 03 and 06 were CDC IVC1.
O_CM.CA9	X78476 U53175	Janssens et al, Braaten et al.	The env sequence was kindly released for publication in the 1995 Human Retroviruses and AIDS database by Dr. Wouter Janssens. The Pol region was published by Janssens et al, 1994, and the p24 region by Braaten et al., 1996.
O_GA.VI686	X78476 X96526	Janssens et al. Delaporte et al.	The VI686 viral sequence spanning the pol region was published by Janssens et al., and the complete Env gp160 was published in a diversity study of HIV isolates from Gabon by Delaporte et al.
O_CM.YBF 22,26,28,32,35, and 37		(F. Simon, P. Mauclere, and S. Saragosti, submitted)	These sequences are derived from a set of individuals from Yaounde and have been submitted for publication. 22 and 26 were CDC stage II, the other four had AIDS. They were isolated in 1996.

Table 1 cont.

Sequence	Accesion	Reference	Note
O_FR.DUR or RUD	X84327		This group O viral sequence was derived from a French woman who had a Cameroonian sexual contact. It was made available in GenBank prior to publication by J. Cohen.
O_CM.MVP2171		(Lutz Gurtler, personal communication)	The complete genome of this sequence has been finished, but it is not yet available at the database. The immunodominant region of gp41 was available, and is included here.

GenBank accession numbers of additional sequences shown in the alignments.

Env sequences

A_U455 M62320
A_KENYA L22943
B_D31 U43096
B_LAI K02013
C_SM145A L22946
C_DJ373A L23065
D_Z2Z6 M22639
D_ELI K03454
E_TN235 L03698
E_CM240X L14572
F_BZ163A L22085
F_93BR020.17 U27401
G_LBV217 U09664
G_92RU131.9 U30312
H_CA13 U09667
H_VI557 L11793
CPZGAB X52154
CPZANT U42720

p24 sequences

CONSENSUS-M
A_KE_K88 L11773
A_ZR_VI57 L11794
B_DE_D31 U43096
B_FR_LAI K02013
C_SO_SM145 L11803
C_ZR_VI313 L71787
D_ZR_ELI K03454
D_ZR_Z2Z6 M22639
F_BR_BZ162 L11751
F_ZR_VI174 L11782
G_ZR_VI191 L11783
H_GA_VI525 L11792
H_ZR_VI557 U09666
CPZGAB X52154
CPZANT U42720

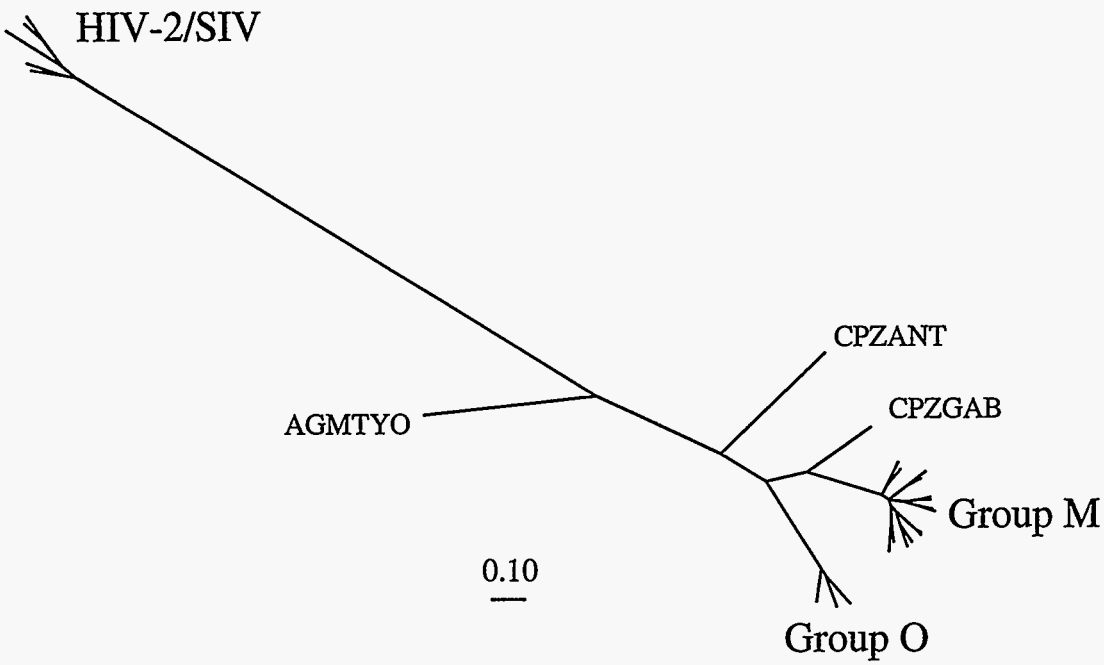


Figure 1. A maximum likelihood tree illustrating the phylogenetic relationship between HIV-1 Group M, HIV-1 Group O, and chimpanzee viral sequences relative to other primate lentiviral sequences. This tree was created using PHYLIP DNAML 4.0 (J. Felsenstein, `http://evolution.genetics.washington.edu/phylip`), with a series of optimization tests performed to select the parameters of three rate categories (1 3 and 9 with frequencies of 0.25, 0.50 and 0.25, respectively) and a transition/transversion ratio of 1.3. The sequences used were the group M isolates subtypes A–G from the alignments in Figure 1 and 2; the full length envelope group O sequences (MVP50, ANT70, VAU); AGMTYO; HIV-2/SIV sequences BEN, D205, ROD, and SIVMNE and SIVSTM; and the chimpanzee viral sequences CPZGAB and CPZANT70. There were 1866 positions spanning gp160 included in the alignment after gapstripping. The taxa included in this comparison were aligned using a Hidden Markov Model (HMMER version 1.8 `http://genome.wustl.edu/eddy/HMMER/main.html`) [Myers & Farmer(1996), Eddy (1995)]. The clade including the HIV-1 and chimpanzee viral sequences is also seen in 100/100 PHYLIP Neighbor Joining bootstrap replicates using a p24 region alignment. The chimpanzee sequences are sometimes linked in phylogenetic reconstructions, and other times are separated as is seen is this tree, with CPZGAB branching with the Group M isolates. The associations of the CPZ sequences depend upon the region of HIV-1 used to construct the tree (data not shown), and may reflect an early recombination event.

sequences of the envelope gene fragments spanning the C2V3 region have been shown to be particularly phylogenetically informative [Leitner (1996)] and to give accurate phylogenetic reconstructions of the M group clades using a variety of phylogenetic tree reconstruction methods [Korber (1994), Leitner(1996)]. Despite this, very little clear phylogenetic structure can be resolved among the group O sequences, in either the region encoding C2V3 of envelope or p24 (analysis was conducted on all sequences available, data not shown). This lack of clearly defined phylogenetic branching patterns in group O could be due to one or a combination of the following three possibilities.

i) An artifact of sample size; accumulation of sequences from new isolates should help us to evaluate this hypothesis.

ii) The length of the sequences in the regions studied not being adequate to reveal phylogenetic associations in group O, although the subtypes are clearly distinguishable using these regions for group M.

iii) A distinct evolutionary history of the two groups, such that the epidemiological history and biology of group M virus yields geographically and phylogenetically distinct subtypes, whereas the group O isolates had a different natural history and simply radiated out from an original ancestral virus [Loussert-Ajaka (1995)]. Interestingly, the first case of AIDS documented in Europe was a Norwegian sailor infected with a group O virus [Jonassen (1997)]. He developed clinical manifestations of HIV infection in 1966, his wife in 1967, and daughter in 1969. All of them died in the 1970s [Froland (1988)].

Geographical distribution: Group O isolates can be identified genetically or serologically. Group O isolates that have been genetically characterized originated in Norway [Jonassen (1997)], Belgium [de Leys (1990)], France [Loussert-Ajaka (1995)], Germany [Hampl (1995)], Spain [Soriano (1996)], the United States [Rayfield (1996)], and, of course, Cameroon [Mauclere (1997)]. Group O infections have been serologically characterized using sera originating in Cameroon and Gabon [Nkengasong (1994)], Equatorial Guinea [Hunt (1996)], Benin [Heyndrickx (1996)], and Kenya [Songok (1996)]. In Cameroon, 7 group O infections were found among 332 sera tested (7/332), Gabon (2/213), Niger (5/1459), Nigeria (2/183), Senegal (1/1283), Chad (2/619), and Togo (1/670) [Peeters (1997)].

It is clear that Cameroon has the highest known prevalence of group O infections. Epidemiology studies conducted in Cameroon indicate a prevalence of HIV-1 group O infection of 2% [Zekeng (1994)], and more recently, Mauclere et al. found prevalence of group O infections ranging from 1% in the northern part of the country, to 6.3% in the capitol; overall, 82 group O infections were found among 2458 HIV-1 samples [Mauclere (1997)]. Among the 19 cases found in France, 17 were from patients originally from Cameroon, and of the two French Caucasian patients, one acknowledged having a Cameroonian sexual partner.

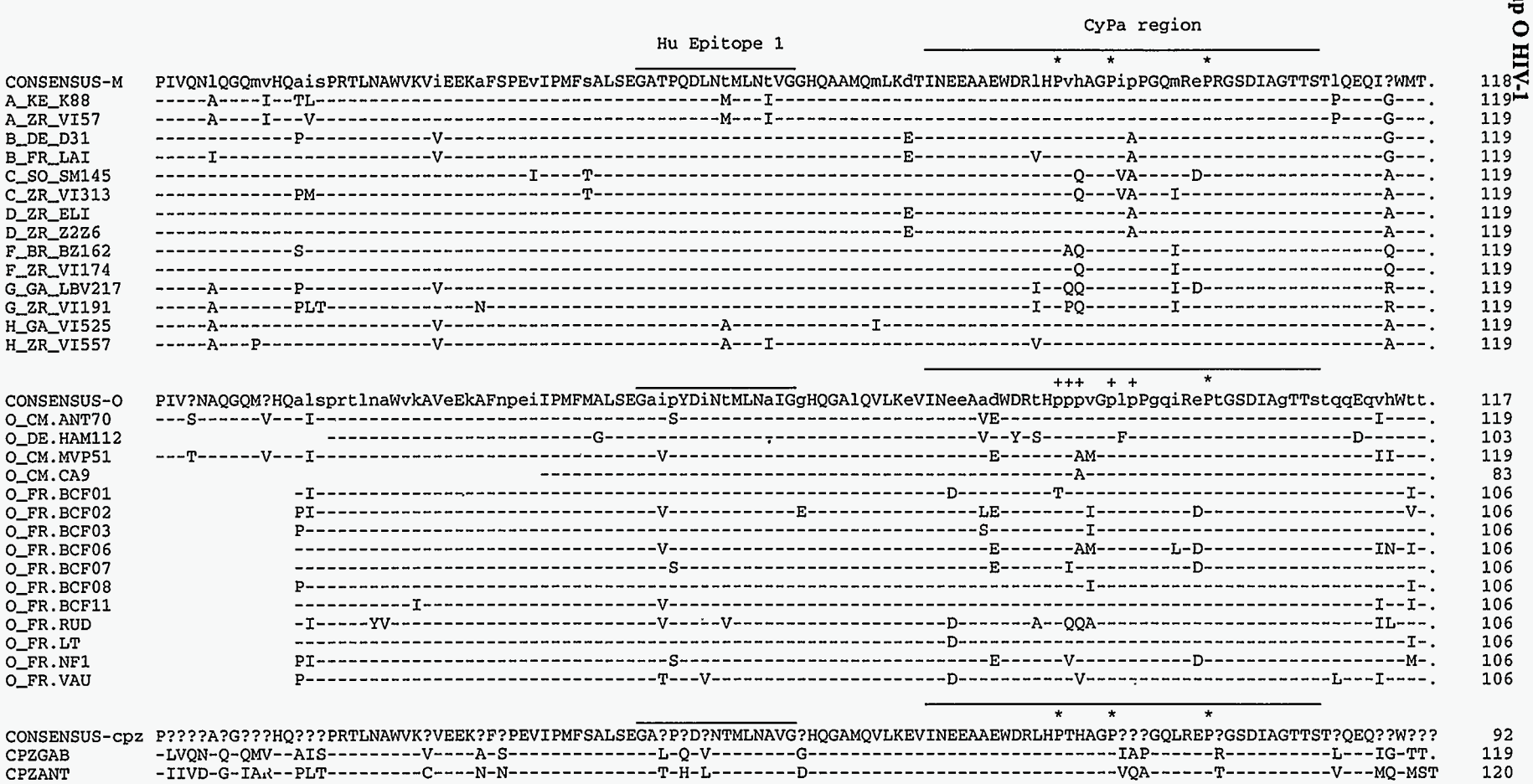
The algorithm developed to characterize group O viruses [Gurtler(1996), Mauclere (1997)] has proven to be very effective. The development of new tools (for example, new peptides used in combination, LIA (Innogenetics)) has allowed the easy serological characterization of group O isolates and the screening of large numbers of samples.

The gag capsid (p24) protein

Background: The gag precursor poly-protein is initially incorporated into the budding viral particle, and is cleaved into the structural proteins of the mature virion by viral protease. The Gag precursor self-assembles into virion like proteins even in the absence of other viral proteins; it binds and packages viral RNA, and interacts with other viral proteins and envelope lipid to assemble the viral particles (for a review of Gag proteins, see [Gorelick & Henderson(1994)]). The capsid antigen (CA) is the p24 protein of HIV-1, and forms the viral core.

The major homology region: The major homology region (MHR) is a part of the capsid protein which is well conserved among lentiviruses of primates, cats, and ungulates [Otteken (1996), Matsuo (1992), Grund (1994), Gorelick & Henderson(1994)]. This region plays a role in viral particle formation and is critical for viral replication. Four highly conserved positions among all retroviruses (human, simian, caprine, bovine, equine, and feline) [Mammano (1994)] are indicated by asterisks in the alignment shown in Figure 2, on the following two pages, and in the consensus MHR alignment (Figure 3). The first three of these positions are critical for HIV-1 particle production, and mutations in forth residue result in aberrant particle size and inappropriate viral cores [Mammano (1994)].

The MHR is conserved among group O viral sequences, as expected, although slightly greater variation is apparent in the group O viruses than the M group viruses. The conserved arginine (R) is present in most retroviral sequences, but it is substituted with a serine (S) in two of the group O viral sequences (data not shown).


Hu Epitope 1
CyPa region
CONSENSUS-M PIVQNlQGQmvHQaisPRTLNAWVKViEEKaFSPEvIPMFsALSEGATPQDLNtMLNtVGGHQAAMQmLKdTINEEAAEWDRlHPvhAGPipPGQmRePRGSDIAGTTSTlQEQI?WMT. 118
A_KE_K88 -----A----I--TL----------M---I----------P----G---. 119
A_ZR_VI57 -----A----I---V----------M---I----------P----G---. 119
B_DE_D31 ----------P----------V----------E----------A----------G---. 119
B_FR_LAI -----I----------V----------E----------V--------A----------G---. 119
C_SO_SM145 ----------I----T----------Q---VA-----D----------A---. 119
C_ZR_VI313 ----------PM----------T----------Q---VA---I----------A---. 119
D_ZR_ELI ----------E----------A----------A---. 119
D_ZR_Z2Z6 ----------E----------A----------A---. 119
F_BR_BZ162 ----------S----------AQ--------I----------Q---. 119
F_ZR_VI174 ----------Q--------I----------Q---. 119
G_GA_LBV217 -----A-------P----------V----------I--QQ--------I-D----------R---. 119
G_ZR_VI191 -----A-------PLT----------N----------I--PQ--------I----------R---. 119
H_GA_VI525 -----A----------V----------A----------I----------A---. 119
H_ZR_VI557 -----A---P----------V----------A---I----------V----------A---. 119
CONSENSUS-O PIV?NAQGQM?HQalsprtlnaWvkAVeEkAFnpeiIPMFMALSEGaipYDiNtMLNaIGgHQGAlQVLKeVINeeAadWDRtHpppvGplpPgqiRePtGSDIAgTTstqqEqvhWtt. 117
O_CM.ANT70 ---S------V---I----------S----------VE----------I----. 119
O_DE.HAM112 ----------G----------V--Y-S-------F----------D------. 103
O_CM.MVP51 ---T------V---I----------V----------E-------AM----------II---. 119
O_CM.CA9 ----------A----------. 83
O_FR.BCF01 -I----------D---------T----------I-. 106
O_FR.BCF02 PI----------V----------E----------LE---------I---------D----------V-. 106
O_FR.BCF03 P----------S---------I----------. 106
O_FR.BCF06 ----------V----------E-------AM-------L-D----------IN-I-. 106
O_FR.BCF07 ----------S----------E------I----------D----------. 106
O_FR.BCF08 P----------I----------I-. 106
O_FR.BCF11 ----------I----------V----------I--I-. 106
O_FR.RUD -I-----YV----------V----V----------D-------A--QQA----------IL---. 106
O_FR.LT ----------D----------I-. 106
O_FR.NF1 PI----------S----------E------V----------D----------M-. 106
O_FR.VAU P----------T---V----------D----------V----------L---I----. 106
CONSENSUS-cpz P????A?G???HQ???PRTLNAWVK?VEEK?F?PEVIPMFSALSEGA?P?D?NTMLNAVG?HQGAMQVLKEVINEEAAEWDRLHPTHAGP???GQLREP?GSDIAGTTST?QEQ??W??? 92
CPZGAB -LVQN-Q-QMV--AIS---------V----A-S----------L-Q-V--------G----------IAP------R----------L---IG-TT. 119
CPZANT -IIVD-G-IAR--PLT---------C----N-N----------T-H-L--------D----------VQA------T----------V---MQ-MST 120

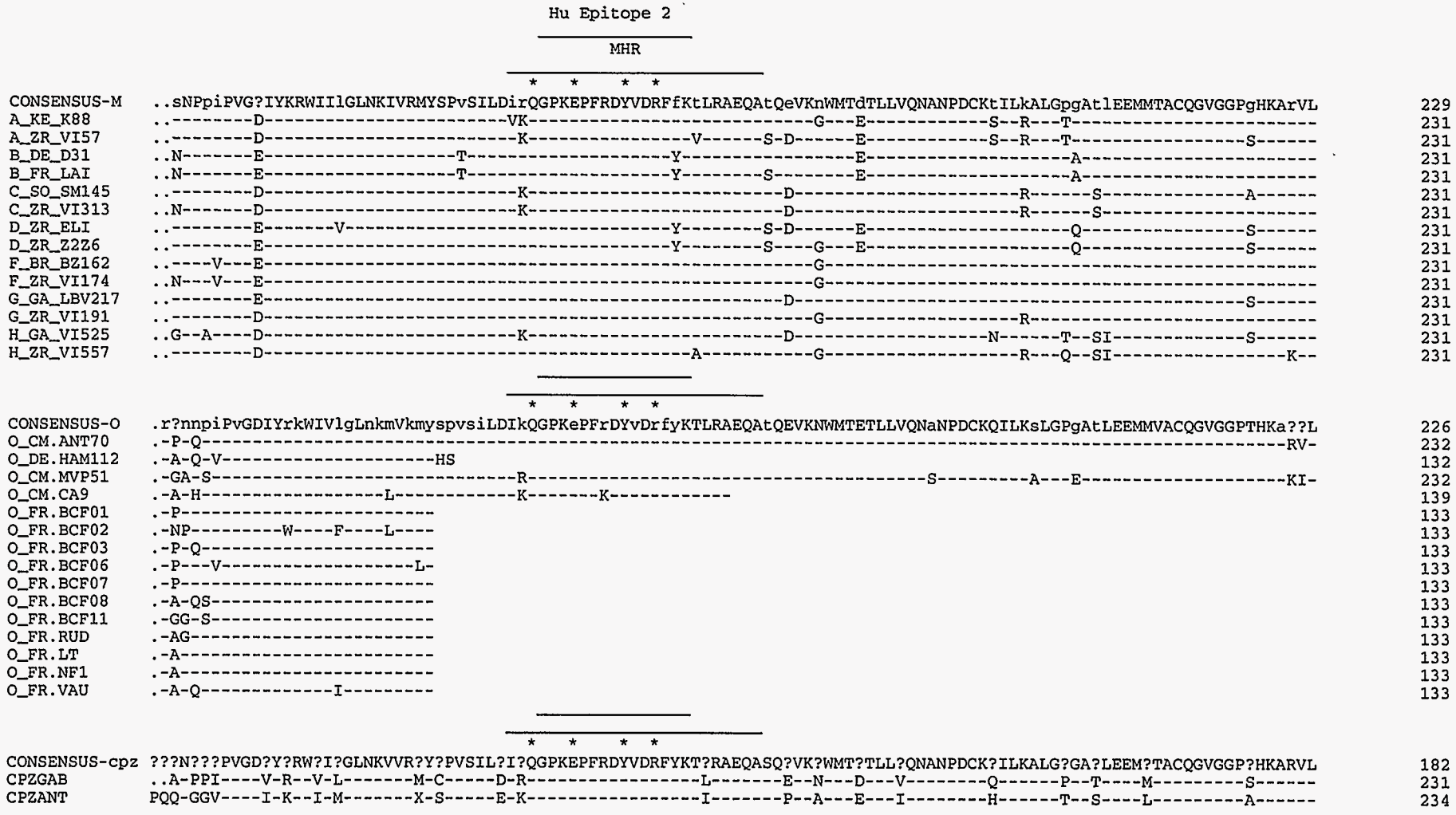


Figure 2. Alignment of the p24 protein. The group M consensus is based on the sequences shown, so each M subtype is equally represented, by two sequences. Upper case letters signify invariant amino acids, and lower case indicate the most common amino acid in a position with some variation. The group O consensus was constructed from all 42 sequences available, but only those published or released are shown. A dash indicates identity with the consensus, a period indicates a insertion made to maintain the alignment. Hu Epitope 1 and 2 are immunodominant linear B cell epitopes that have been identified in group M [Janvier (1996)], and are highly conserved between groups M and O. The MHR is the major homology region and the CyPa region is the cyclophilin A binding domain. Asterisks in the MHR mark highly conserved, functionally important residues, and in the CyPa region mark prolines critical for group M viral replication. The plus signs over the O group consensus indicate prolines that are common among O group sequences, but not perfectly conserved.

```
                          *   *     *  *
M group consensus:    irQGPKEPFRDYVDRFfKtLRAEQA
group O consensus:    -k----e--r--v-rfy--------
     cpz consensus:   -?--------------Y--?-----
```

Figure 3. A comparison of the consensus sequences across the MHR of gag p24. Highly conserved amino acids essential for group viral particle formation are marked with asterisks.

There is an immunodominant region within the MHR that can stimulate both human and murine antibodies (see [Korber (1996)] for a summary). One of the immunodominant linear epitopes defined for HIV-1 M group viruses is located within this region, defined as: GPKEPFRDYVDRFYK, (called Hu Epitope 2 in Figure 2) [Janvier (1996)]. Only a single amino acid differs in this epitope between the group O consensus the M group consensuses. The tyrosine (Y) in the second to last position of the epitope matches the consensus of group O sequences, and is found in all B and D clade sequences, as well as among the Thai A subtypes. But in all other clades in the M group, this position is most commonly occupied by a phenylalanine (F). It is not clear whether this substitution would influence the antigenic specificity of the peptide, though it is very likely that the antigenic peptide studied by Janvier et al. would react with sera from individuals infected with group O strains of HIV-1, because the group M subtype B derived peptide perfectly matches the group O consensus. 9/20 sera from HIV-1 infected individuals reacted with this peptide [Janvier (1996)]. Among HIV-2 and SIV sequences, the immunogenic region is conserved except for two substitutions: the central RD in positions six and seven become QS (see the Gag protein alignments in this compendium). These substitutions may not be critical for binding, as murine monoclonal antibodies that bind to this region have been shown to be cross-reactive with HIV-1, HIV-2, and SIVmac and SIVagm strains [Matsuo (1992), Robert-Hebmann (1992), Niedrig (1989)]. Therefore, this immunodominant domain could potentially be useful for serological identification of all primate lentiviral infections, by sequence analogy, including those of group O.

Cyclophilin A binding region: Cyclophilin A (CyPa) is a human protein that binds to a proline rich region in the HIV-1 p55 Gag precursor protein, and is incorporated into group M HIV-1 viral particles, (). Group M HIV-1 viral particles that lack cyclophilin A are not infectious, in contrast to other retroviruses [Luban (1994), Thali (1994), Franke (1994)]. Cyclophilins are a family of proteins that catalyze protein folding and are protective against heat shock [Fisher & Schmid(1990), Gething & Sambrook(1992), Sykes (1993)]. In contrast to HIV M group virus and the SIVcpz GAB strain, and the group O viruses do not have to bind and incorporate CyPa to produce infectious virions; this is one of the first major biological differences between group M and group O to be documented [Braaten (1996)].

Three of the prolines in the highly conserved CyPa binding region of p24 are required for the production of infectious virions in M group strains. These prolines are marked with an asterisk in Figures 2 and 4. The proline at position 222 (AGP) was required for CyPa binding and incorporation into virions, and was required for replication. The proline at position 225 (PGQ) was not essential for replication, and so is not marked. Mutations in prolines 217 (lHP) and 231 (PRG) decreased particle yield and the mutant virus did not replicate [Franke (1994)]. Proline 222 was highly, but not perfectly, conserved among group O viruses, in spite of the lack of a requirement for cyclophilin A binding. Proline 231 was perfectly preserved among group M, group O, and SIVcpz. Proline 217 is perfectly preserved among group M and SIVcpz viruses, but varies somewhat in group O viruses. Two additional prolines are present in group O viruses proximal to position 217 (Figure 2).

There are many murine monoclonal antibodies that bind to the CyPa binding domain [Korber (1996)]. One particular epitope that has been defined by peptide reactivity (ETINEEAAEWDRVHP) was cross-reactive with sera from HIV-1, HIV-2 and SIV infections [Niedrig (1988)]. While there are frequent substitutions among group O isolates in this region, most are conservative aspartic acid (D) to glutamic acid (E) substitutions, and there are some O strains that are identical to the M consensus sequence in this region. Thus it is likely that group O sera would react with this peptide.

```
                                  *    *         *
M group consensus:    INEEAAEWDRlHPvhAGPipPGQmRePRGSDIAGTTS
group O consensus:    --ee-ad---t-pppv-plp-gqi-e-t-----g--s
    cpz consensus:    ----------L--TH---???--L-E-?---------
```

Figure 4. A comparison of the consensus sequences across the Cyclophilin A binding region in gag p24. Highly conserved prolines essential for group M viral replication are marked with asterisks.

Additional immunogenic domains in p24: A peptide covering positions 178-192 (HIV-1 IIIB) was recognized by sera from 8/20 HIV-1 positive people (Hu Epitope 1 in Figure 2) [Janvier (1996)]. The linear immunodominant antigenic regions (Hu Epitope 1 and 2) in p24 group M viruses are conserved in group O as well, particularly human epitope 2, suggesting that they may provide broadly cross-reactive antigenic peptides recognized by sera of individuals with O and M group infections.

Through studies involving M group strains, HIV-1 epitopes have been defined for both helper and cytotoxic T cells across the p24 protein. The p24 T-cell epitopes that have been defined in M group tend to cluster, some near the C-terminal region, and some crossing the CyPa binding site and the MHR [Korber (1996)]. Because of the conservation of p24, it is highly probably that some cross-reactivity between M and group O would be observed. For example, there is an HLA-A25 restricted CTL epitope, (ETINEEAAEW) that has been defined using CTL clones from long term survivors [van Baalen (1996)]. This epitope is conserved in viruses of the B and D subtypes, but the form DTINEEAAEW is commonly in other clades, and this variant was not recognized by CTL specific for the index peptide. While group O sequences show some variation in this epitope, some are very similar to the reactive M group epitope, and may be cross-reactive: (*e.g.* MVP51: EVINEEAAEW).

The C2V3 Region

Variation and immunogenicity of the V3 loop: The variation in the V3 loop has been the focus of extensive international research efforts, originally because it was noted in the late 1980s that antibodies directed against the tip of V3 loop could potently neutralize laboratory strains of HIV-1 in a type-specific manner [Javaherian (1989), Palker (1988), Rusche (1988)]. More recently, anti-V3 monoclonal antibodies have been shown in general to be potent neutralizing agents when directed against HIV-1 laboratory-adapted strains, but far less so when directed against primary isolates. This presumably is due to distinct conformations of envelope, such that the V3 loop of primary isolates is inaccessible (see [Moore & Ho(1995), Poignard (1996)] for review). Most anti-V3 monoclonal antibodies are type-specific, and monoclonals directed against one strain can be completely unreactive even with closely related viral strains, due to the hypervariable nature and high rate of amino acid substitution in the immunogenic tip of the V3 loop [Korber (1996)]. Other neutralizing epitopes, both continuous and discontinuous, have been identified in envelope [Poignard (1996), Moore & Ho(1995)]. Many neutralizing antibodies (but not all, *e. g.* [Trkola (1995)]) share the property of reduced neutralization of primary isolates relative to laboratory adapted strains [Moore & Ho(1995)].

The turn at the tip of the loop is the focus of most anti-V3 neutralizing antibodies, and is most often found to be GPGR or GPGQ among M group viruses [Foley (1996)]. These two motifs are not found among group O strains. With the exception of some common amino acids flanking the cysteines at the base of the V3 loop, group O and group M V3 loops are very different from each other. Helper and cytotoxic T cell epitopes, as well as antibody binding sites, have been identified within the V3 loop of M group viruses. Because group O and M viruses are highly variable in the C2V3 region within a group, and because the extent of the substitutions found between group O and M is so great, it is highly unlikely that there would be antigenic O and M cross-reactivity for either B or T cell epitopes in the V3 loop.

Phenotypic determinants in the V3 loop: The V3 loop has been shown to have a critical functional role in determining the phenotype of the virus [Hwang (1991), Chesebro (1992), de Jong (1992)]. Positively charged amino acids in certain positions in the V3 loop are correlated with a syncytium-inducing (SI), T-cell-tropic viral phenotype among group M subtype B viruses [Fouchier (1992)] (the two most critical positions are indicated in Figure 5). The observations concerning these two positions, and a correlation between net positive charge on the loop and rapid high and slow low

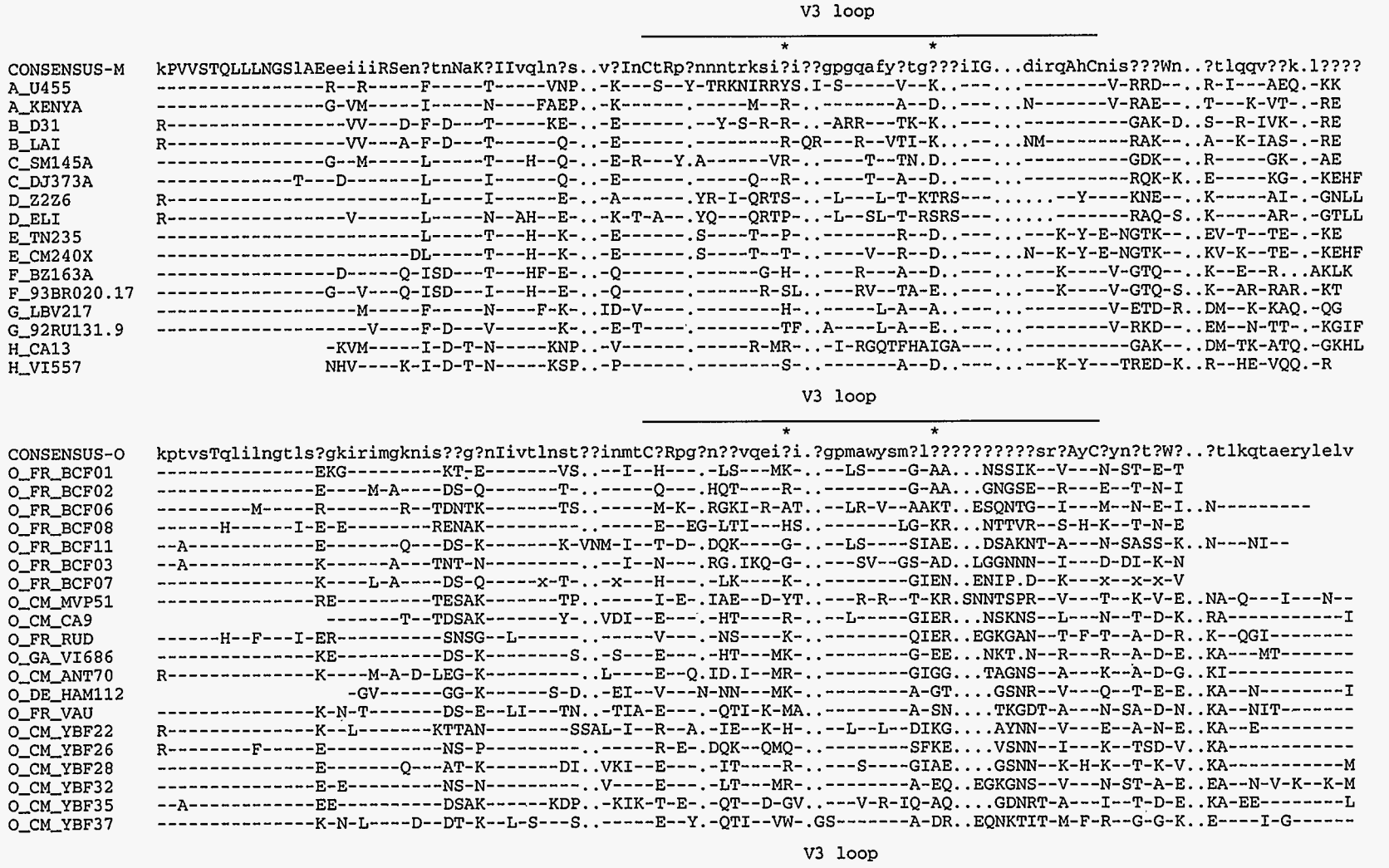


Figure 5. Alignment of the C2V3 region of gp120. The boundaries of the V3 loop are shown, and the positions that are associated with viral phenotype (SI phenotype and positive charge) are marked with asterisks. 45 group O sequences were used to generate the consensus, but only those published or released are shown.

phenotypes, were extended beyond subtype B to other group M subtypes [de Wolf (1994), Rubsamen-Waigmann (1994)]. Recently, a correlation between positive charge and phenotype was shown for the group O viruses [de Jong (1996)]. Because of the length variation between different strains and the diversity found in this region, it is difficult to be certain the alignment of V3 loop shown in Figure 5 is biologically appropriate. Despite this, in a best estimate the two critical sites in the V3 loop associated with group M SI/NSI phenotype switching display a similar pattern of variation among group O strains and group M sequences. Both are highly variable positions, and the first position alternates between neutral and positively charged amino acids, and the second position alternates between negative, neutral and positively charged amino acids.

Recently it has been shown that chemokine second receptor usage for HIV entry into CD4-positive immune cells is associated with the phenotype of the virus, and not with the genetically defined HIV-1 group or subtype [Zhang (1996), Choe (1996)]. CXCR4 (fusin) usage is associated with viruses of with T-cell tropic, SI phenotypes; CCR5 usage is associated with a macrophage tropic, non-SI (NSI) phenotype; and some SI viruses are able to use both CXCR4 and CCR5 [Zhang (1996)]. The group O viruses, similar to the M group, have both patterns of second receptor usage [Zhang (1996)]. The two group O isolates for which second receptor usage has been determined are CA9, with an NSI phenotype and second receptor usage restricted to CCR5; and MVP5180, with an SI phenotype and the ability to use both CCR5 and CXCR4. The pattern of positive charge substitutions in the V3 loop, at least in these two group O isolates, is consistent with what has been observed in the M group. The SI isolate MVP5180 has greater positive charge overall, and in particular, in the second of the two sites associated with the SI phenotype in the M group is positively charged in O_CM_MVP51 and negatively charged in O_CM_CA9 (Figure 6). The C-C chemokines RANTES, MIP-1α, and MIP-1β produced by CD8+ cells, can inhibit HIV-1 infection of primary and macrophage-tropic group M HIV-1 isolates, whereas T-cell tropic isolates tend to be resistant to the C-C chemokine suppressive effects [Cocchi (1995)]. The V3 loop was shown to be a critical region for determining the susceptibility to chemokine-mediated supression using chimeric gp120 proteins [Cocchi (1996)]. It will be interesting to see if SI group O viruses, like MVP5180, will have a similar pattern of resistance.

```
                          *             *
                              + +     ++
O_CM_MVP51    CIREGIAEVQDIYTGPMRWRSMTLKRSNNTSPRSRVAYC
  O_CM_CA9    CERPGNHTVQEIRIGPLAWYSMGIERNSKNSSR--LAYC
```

Figure 6) The V3 loop sequences of SI strain MVP51 and NSI strain CA9. The positions that have been associated with viral phenotype are marked with asterisks. The positively charged amino acids arginine (R) and lysine (K) near the tip of the loop V3 found in SI isolate MVP51 are marked with a plus (+). The dashes (-) indicated insertions made to maintain the alignment.

Immunodominant domain in gp41

The immunodominant region in gp41 is conserved between different group M strains, and is reactive with sera from individuals infected with a variety of group M subtypes [Engelbrecht (1994), Petrov (1990), Shafferman (1989), Gnann (1987)]. Although only a limited number of sequences are avalailable in the database across this region, because of its importance for diagnostics, a comparison of the and group M and O consensus sequences in this region is shown in Figure 7.

Acknowledgments

With thanks to Lutz Gurtler, Sushil Devare, Francois Simon, Philippe Mauclere, J. Cohen, Marie Laure Chaix, Guido Van der Groen and Wouter Janssens for their history of graciously providing sequences to the Los Alamos Database prior to publication of the primary articles introducing the sequences, for sharing their knowledge concerning group O viruses, and for contributions of sequences used to build the consensuses provided in this manuscript.

Figure 7. Alignment of amino acids from gp41 spanning the immunodominant domain of gp41. This alignment starts at 50 amino acids from the start of gp41 and continues for the next 100 amino acids. The M consensus is based on the set of sequences shown. The group O consensus is based on the 9 sequences currently available at the Los Alamos database covering this region; the five that have been published or made public are shown in the alignment.

References

[Braaten (1996)] D. Braaten, E. K. Franke, & J. Luban. Cyclophilin A is required for the replication of the group M Human Immunodeficiency virus type 1 (HIV-1) and simian Immunodeficiency virus SIVcpzGAB but not group O HIV-1 or other primate immunodeficiency viruses. *J Virol* **70**:4420–4427, 1996.

[Charneau (1994)] P. Charneau, A. M. Borman, C. Quillent, D. Guetard, S. Chamaret, J. Cohen, G. Remy, L. Montagnier, & F. Clavel. Isolation and envelope sequence of a highly divergent HIV-1 isolate: definition of a new HIV-1 group. *Virology* **205**:247–253, 1994.

[Chesebro (1992)] B. Chesebro, K. Wehrly, J. Nishio, & S. Perryman. Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell tropic isolates: definition of critical amino acids involved in cell tropism. *J Virol* **66**:6547–6554, 1992.

[Choe (1996)] H. Choe, M. Farzan, Y. Sun, N. Sullivan, B. Rollins, P. D. Ponath, L. Wu, C. R. Mackay, G. LaRosa, W. Newman, N. Gerard, C. Gerard, & J. Sodroski. The β-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell* **85**:113–1148, 1996.

[Cocchi (1995)] F. Cocchi, A. L. DeVico, A. Garzino-Demo, S. K. Arya, R. C. Gallo, & P. Lusso. Identification of RANTES, MIP-1α and MIP-1β as the major HIV-suppressive factors produced by CD8+ T cells. *Science* **270**:1811–1815, 1995.

[Cocchi (1996)] F. Cocchi, A. L. DeVico, A. Garzino-Demo, A. Cara, R. C. Gallo, & P. Lusso. The V3 domain of the HIV-1 gp120 envelope glycoprotein is critical for chemokine-mediated blockage of infection. *Nature Medicine* **270**:1811–1815, 1996.

[Cohen (1995)] J. Cohen, J. de Saint Martin, S. Chamaret, D. Guetard, L. Montagnier, T. Tabary, F. Philibert, & F. Bougy. The HIV-O strain DUR illustrates divergences and homologies in the HIV-1 O group. *First HIV-O Symposium* 1995. 3rd-4th March 1995. Munchen, Germany.

[de Jong (1996)] J. de Jong, F. Simon, G. van der Groen, E. Baan, S. Saragosti, F. Brun-Vezinet, & J. Goudsmit. V3 loop sequence analysis of seven HIV type 1 group O isolates phenotyped in peripheral blood mononuclear cells and MT-2 cells. *AIDS Res and Human Retroviruses* **12**:1503–1507, 1996.

[de Jong (1992)] J. J. de Jong, J. Goudsmit, W. Keulen, B. Klaver, W. J. A. Krone, M. Tersmette, & A. de Ronde. Human immunodeficiency virus type 1 clones chimeric for the envelope V3 domain differ in syncytium formation and replication capacity. *J Virol* **66**:757–765, 1992.

[de Leys (1990)] R. de Leys, B. Vanderborght, M. vanden Haesevelde, L. Heyndrickx, A. van Geel, C. Wauters, R. Bernaerts, E. Saman, P. Nijs, B. Willems, H. Taelman, G. van der Groen, P. Piot, T. Tersmette, J. G. Huisman, & H. van Heuverswyn. Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. *J Virol* **64**:1207–1216, 1990.

[de Wolf (1994)] F. de Wolf, E. Hogervorst, J. Goudsmit, E. Fenyo, H. Rubsamen-Waigmann, & et al. Syncytium inducing (SI) and non-syncytium inducing (NSI capacity of human immunodeficiency virus type 1 (HIV-1) subtypes other than B: phenotypic and genotypic characteristics. *AIDS Res and Human Retroviruses* **10**:1387–1400, 1994.

[Descamps (1995)] D. Descamps, G. Collin, I. Loussert-Ajaka, S. Saragosti, F. Simon, & F. Brun-Vezinet. HIV-1 group O sensitivity to antiretroviral drugs. *AIDS* **9**:977–978, 1995.

[Eddy (1995)] S. Eddy, G. Mitchison, & R. Durbin. Maximum discrimination hidden markov models of sequence consensus. *J Comp Biol* **2**:9–23, 1995.

[Engelbrecht (1994)] S. Engelbrecht, G. J. de Jager, & E. J. van Rensburg. Evaluation of commercially available assays for antibodies to HIV-1 in serum obtained from South African patients infected with HIV-1 subtypes. *J Med Virol* **44**:223–228, 1994.

[Fisher & Schmid(1990)] G. Fisher & F. X. Schmid. The mechanism of protein folding: implications of *in vitro* refolding models for *de novo* protein folding and translocation in the cell. *Biochemisty* **29**:2205–2212, 1990.

[Foley (1996)] B. T. Foley, J. C. Blouin, & E. A. Guzman. Global Variation in the HIV-1 V3 Region pages part III-77, 1996. Published in this compendium.

[Fouchier (1992)] R. A. M. Fouchier, M. Groenink, N. A. Kootstra, M. Tersmette, H. G. Huisman, F. Miedema, & H. Schuitemaker. Phenotype-associated sequence variation in the third variable region of the human immunodeficiency virus type 1 gp120 molecule. *J Virol* **66**:3183–3187, 1992.

[Franke (1994)] E. K. Franke, H. E. H. Yuan, & J. Luban. Specific incorporation of cyclophilin A in HIV-1 virions. *Nature* **372**:359–362, 1994.

[Froland (1988)] S. S. Froland, P. Jenum, C. F. Lindboe, K. W. Wefring, P. J. Linnestad, & T. Bohmer. HIV-1 infection in norwegian family before 1970. *Lancet* **331**:1344–1345, 1988.

[Gething & Sambrook(1992)] M. J. Gething & J. Sambrook. Protein folding in the cell. *Nature* **355**:33–45, 1992.

[Gnann (1987)] J. W. Gnann, P. L. Schwimmbeck, J. A. Nelson, A. B. Truax, & M. B. Oldstone. Diagnosis of AIDS by using a 12-amino acid peptide representing an immunodominant epitope of the human immunodeficiency virus. *J Infect Dis* **156**:261–267, 1987.

[Gorelick & Henderson(1994)] R. J. Gorelick & L. E. Henderson. *Gag Proteins*. Los Alamos National Laboratory: Theoretical Biology and Biophysics, Los Alamos, NM, 1994. Editors: G. Myers and B. Korber and K.-T. Jeang and L. Henderson and S. Wain-Hobson and G. N. Pavlakis.

[Grund (1994)] C. H. Grund, E. R. Lechman, C. J. Issel, R. C. Montelaro, & K. E. Rushlow. Lentivirus cross-reactive determinants present in the capsid protein of equine infectious anaemia virus. *J Gen Virol* **75**:657–662, 1994.

[Gurtler(1996)] L. Gurtler. Difficulties and strategies of HIV diagnosis. *Lancet* **348**:176–179, 1996.

[Gurtler (1994)] L. G. Gurtler, P. H. Hauser, J. Eberle, A. von Brunn, S. Knapp, L. Zekeng, J. M. Tsague, & L. Kaptue. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. *J Virol* **68**:1581–1585, 1994.

[Hampl (1995)] H. Hampl, D. Sawitzky, M. Stoffler-Meilicke, A. Groh, M. Schmitt, J. Eberle, & L. Gurtler. First case of HIV subtype O infection in Germany. *Infection* **23**:369–370, 1995.

[Heyndrickx (1996)] L. Heyndrickx, M. Alary, W. Janssens, N. Davo, & G. van der Groen. HIV-1 group O and group M dual infection in Benin. *Lancet* **347**:902–903, 1996.

[Hunt (1996)] J. C. Hunt, A. M. Golden, A. Vallari, & et al. Molecular and serologic characterization of four HIV-1 group O from Equatorial Guinea. *AIDS Res Hum Retroviruses* **11 Suppl. 1**:S144, 1996.

[Hwang (1991)] S. S. Hwang, T. J. Boyle, H. K. Lyerly, & B. R. Cullen. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science* **253**:71–74, 1991.

[Janvier (1996)] B. Janvier, J. J. Lasarte, P. Sarobe, J. Hoebeke, A. Baillou-Beaufils, F. Borras-Cuesta, & F. Barin. B cell epitopes of HIV type 1 p24 capsid protein: a reassessment. *AIDS Res. Hum. Retroviruses* **12**:519–525, 1996.

[Javaherian (1989)] K. Javaherian, A. J. Langlois, C. McDanal, K. L. Ross, L. I. Eckler, C. L. Jellis, A. T. Profy, J. R. Rusche, D. P. Bolognesi, S. D. Putney, & T. J. Matthews. Principal neutralizing domain of the human immunodeficiency virus type 1 envelope protein. *Proc Natl Acad Sci USA* **86**:6768–6772, 1989.

[Jonassen (1997)] T. Jonassen, K. Stene-Johansen, E. S. Berg, O. Hungnes, C. F. Lindboe, & B. Grinde. HIV-1 group O infection in Norwegian patients in the 1960s. *Submitted, personal communication, Dr. Bjorn Grinde* 1997.

[Korber (1994)] B. T. M. Korber, K. MacInnes, R. Smith, & G. Myers. Mutational trends in V3 loop protein sequences observed in different genetic lineages of human immunodeficiency virus type-1. *J Virol* **68**:6730–6744, 1994.

[Korber (1996)] E. Korber, J. Moore, P. D'Souza, C. Brander, B. Walker, R. Koup, B. Haynes, & G. Myers. Hiv molecular immunology database 1996 **II**:Parts I and III, 1996. Publisher: Los Alamos National Laboratory Theoretical Biology and Biophysics, website http://hiv-web.lanl.gov/immuno/.

[Leitner(1996)] T. Leitner. Genetic Subtypes of HIV-1 pages III-28, 1996. Published in this compendium.

[Leitner (1996)] T. Leitner, D. Escanilla, C. Franzen, M. Uhlen, & J. Albert. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proc Natl Acad Sci USA* **93**:10864–10869, 1996.

[Loussert-Ajaka (1995)] I. Loussert-Ajaka, M. Chaix, B. Korber, F. Letourneur, E. Gomas, T. Ly, F. Brun-Vezinet, F. Simon, & S. Saragosti. The variability of HIV type 1 group O strains isolated from Cameroonian patients living in France. *J Virol* **69**:5640–49, 1995.

[Luban (1994)] J. Luban, K. A. Bossolt, E. K. Franke, G. V. Kalpana, & S. P. Goff. Human immunodeficiency virus type 1 gag protein binds to cyclophilins A and B. *Cell* **73**:1067–1078, 1994.

[Mammano (1994)] F. Mammano, A. Ohagen, S. Hoglund, & H. G. Gottlinger. Role of the major homology region of human immunodeficiency virus type 1 in virion morphogenesis. *J Virol* **68**:4827–4936, 1994.

[Matsuo (1992)] K. Matsuo, Y. Nishino, T. Kimura, R. Yamaguchi, A. Yamazaki, T. Mikami, & K. Ikuta. Highly conserved epitope domain in major core protein p24 is structurally similar among human, simian and feline immunodeficiency viruses. *J Gen Virol* **73**:2445–2450, 1992.

[Mauclere (1997)] P. Mauclere, I. Loussert-Ajaka, F. Damond, P. Fagot, S. Souquieres, M. Mony Lobe, F.-X. Mbopi Keou, F. Barre-Sinoussi, S. Saragosti, F. Brun-Vezinet, & F. Simon. Serological and virological characterization of HIV-1 group O infection in Cameroon **in press**, 1997.

[Moore & Ho(1995)] J. P. Moore & D. D. Ho. HIV-1 neutralization: the consequences of adaptation to growth on transformed T cells. *AIDS* **9 suppl A**:S117–S136, 1995.

[Myers & Farmer(1996)] G. Myers & A. Farmer. HIV alignments, database searches and structure prediction pages part III-64, 1996. Published in this compendium.

[Niedrig (1989)] M. Niedrig, J. Hinkula, W. Weigelt, J. L'Age-Stehr, G. Pauli, J. Rosen, & B. Wahren. Epitope mapping of monoclonal antibodies against human immunodeficiency virus type 1 structural proteins by using peptides. *J Virol* **63**:3525–3528, 1989.

[Niedrig (1988)] M. Niedrig, J. P. Rabanus, J. L'age Stehr, H. R. Gelderblom, & G. Pauli. Monoclonal atnibodies directed against human immunodeficiency virus (HIV) gag proteins with specificity for conserved epitopes in HIV-1, HIV-2 and simian immunodeficincy virus. *J Gen Virol* **69**:2109–2114, 1988.

[Nkengasong (1994)] J. N. Nkengasong, W. Janssens, L. Heyndrickx, K. Fransen, P. M. Ndumbe, J. Motte, A. Leonaers, M. Ngolle, J. Ayuk, P. Piot, & G. van der Groen. Genotypic subtypes in HIV-1 in Cameroon. *AIDS* **8**:1405–1412, 1994.

[Otteken (1996)] A. Otteken, S. Nick, & W. B. et al. Identification of a gag protein epitope conserved among all four groups of primate immunodeficiency viruses by using monoclonal antibodies. *J Gen Virol* **12**:519–525, 1996.

[Palker (1988)] T. J. Palker, M. E. Clark, A. J. Langlois, T. J. Matthews, K. J. Weinhold, R. R. Randall, D. P. Bolognesi, & B. F. Haynes. Type-specific neutralization of the human immunodeficiency virus with antibodies to env-encoded synthetic peptides. *Proc Natl Acad Sci USA* **85**:1932–1936, 1988.

[Peeters (1997)] M. Peeters, A. Gueye, S. M'Boup, F. Bibollet-Ruche, E. Ekaza, C. Mulanga, R. Rasmata, R. Gandji, P. Mpele, G. Dibanga, B. Koumare, M. Saidou, E. Esu-Williams, J. P. Lombart, W. Badombena, N. Luo, M. vanden Haesevelde, & E. Delaporte 1997. in press.

[Petrov (1990)] R. V. Petrov, R. M. Khaitov, I. G. Sidorovich, S. P. Pavlikov, I. A. Nikolaeva, M. E. Ivachenko, S. M. Andreev, & L. Y. U. Sklyarov. The use of synthetic peptides in the diagnosis of HIV infections. *Biomed Sci* **1**:239–244, 1990.

[Poignard (1996)] P. Poignard, P. J. Klasse, & Q. J. Sattentau. Antibody neutralizatin of HIV-1. *Immunol Today* **17**:239–246, 1996.

[Rayfield (1996)] M. A. Rayfield, P. Sullivan, C. I. B. C.I, L. Britvan, & et al. HIV-1 group O identified for the first time in the United States. *Moratility and Morbidity Weekly Report (MMWR)* **45**:561–564, 1996.

[Robert-Hebmann (1992)] V. Robert-Hebmann, S. Emiliani, M. Resnicoff, F. Jean, & C. Devaux. Subtyping of human immunodeficiency virus isolates with a panel of monoclonal antibodies: identification of conserved and divergent epitopes on p17 and p25 core proteins. *Mol Immunol* **29**:1175–1183, 1992. OTE: Medline: 92408665.

[Rubsamen-Waigmann (1994)] H. Rubsamen-Waigmann, H. von Briesen, H. Holmes, A. Bjorndal, B. Korber, R. Esser, S. Ranjbar, P. Tomlinson, B. Galvao-Castro, E. Karita, S. Sempala, C. Wasi, S. Osmanov, E. M. Fenyo, & the WHO network for HIV isolation and characterization. Standard conditions of virus isolation reveal biological variability of HIV type 1 in different regions of the world. *AIDS Res and Human Retroviruses* **11**:1401–1408, 1994.

[Rusche (1988)] J. R. Rusche, K. Javaherian, C. McDanal, J. Petro, D. L. Lynn, R. Grimaila, A. J. Langlois, R. C. Gallo, L. O. Arthur, P. J. Fischinger, D. P. Bolognesi, S. D. Putney, & T. J. Matthews. Antibodies that inhibit fusion of human immunodeficiency virus-infected cells bind a 24-amino acid sequence of the viral envelope, gp120. *Proc Natl Acad Sci USA* **85**:3198–3202, 1988.

[Shafferman (1989)] A. Shafferman, J. Lennox, H. Grosfeld, J. Sadoff, R. R. Redfield, & D. S. Burke. Patterns of antibody recognition of selected conserved amino acid sequences from the HIV envelope in sera from different stages of HIV infection. *AIDS Res Hum Retroviruses* **5**:33–39, 1989.

[Simon (1994)] F. Simon, T. D. Ly, A. Baillou-Beaufils, V. Fauveau, J. De Saint-Martin, I.Loussert-Ajaka, M. L. Chaix, S. Saragosti, A. M. Courouce, D. Ingrand, C. Janot, & F. Brun-Vezinet. Sensitivity of screening kits for anti-HIV-1 subtype O antibodies. *AIDS* **8**:1628–1629, 1994.

[Songok (1996)] E. M. Songok, D. K. Libondo, M. C. Rotich, S. A. Oogo, & P. M. Tukei. Surveillance for HIV-1 subtypes O and M in Kenya. *Lancet* **347**:1700, 1996.

[Soriano (1996)] V. Soriano, M. Gutteriez, G. Garcia-Lerma, & et al. First case of group O infection in Spain. *Vox Sang* **71**:66, 1996.

[Sykes (1993)] K. Sykes, M. Gething, & J. Sambrook. Proline isomerases function during heat shock. *Proc Natl Acad Sci USA* **90**:5853–5857, 1993.

[Thali (1994)] M. Thali, A. A. Bukovsky, E. Kondo, B. Rosenwirth, C. T. Walsh, J. Sodroski, & H. G. Gottlinger. Specific association of cyclophilin a with human immunodeficiency virus type 1 virions. *Nature* **372**:363–365, 1994.

[Trkola (1995)] A. Trkola, A. B. Pomales, H. Yuan, B. Korber, P. J. Maddon, G. P. Allaway, H. Katinger, C. F. B. III, D. R. Burton, D. D. Ho, & J. P. Moore. Cross-clade neutralization of primary isolates of human immunodeficiency virus type 1 by human monoclonal antibodies and tetrameric CD4-IgG. *J Virol* **69**:6609–6617, 1995.

[van Baalen (1996)] C. A. van Baalen, M. R. Klein, R. C. Huisman, M. E. Dings, S. R. Kerkhof Garde, A. M. Geretti, R. Gruters, C. A. van Els, F. Miedema, & A. D. Osterhaus. Fine-specificity of cytotoxic T lymphocytes which recognize conserved epitopes of the gag protein of human immunodeficiency virus type 1. *J Gen Virol* **77**:1659–1665, 1996.

[vanden Haesevelde (1994)] M. vanden Haesevelde, J. Decourt, R. De Leys, B. Vanderborght, G. van der Groen, H. van Heuverswijn, & E. Saman. Genomic cloning and complete sequence analysis of a highly divergent African human immunodeficiency virus isolate. *J Virol* **68**:1586–1596, 1994.

[Zekeng (1994)] L. Zekeng, L. Gurtler, E. Afane Ze, A. Sam-Abbenyi, G. Mbouni-Essomba, E. Mpoudi-Ngolle, M. Monny-Lobe, J. B. Tapko, & L. Kaptue. Prevalence of HIV-1 subtype O infection in Cameroon: preliminary results. *AIDS* **8**:1626–1628, 1994.

[Zhang (1996)] L. Zhang, Y. Huang, T. He, Y. Cao, & D. D. Ho. HIV-1 subtype and second-receptor. *Nature* **383**:768, 1996.

Retroviral G → A Hypermutation

Simon Wain-Hobson

Institut Pasteur, 28 rue du Dr. Roux, 75724 Paris, Cedex 15, France

A G → A hypermutant is a term given to retroviral proviruses bearing an inordinate number of G → A transitions, typical examples being given in Figure 1. Although two cases were unwittingly described for HIV-1 (Goodenow et al., 1989), the phenomenon was first described for spleen necrosis virus (SNV) (Pathak and Temin, 1990). More examples followed for HIV-1 (Vartanian et al., 1991; Vartanian et al., 1994), SIV (Johnson et al., 1991; Pelletier et al., 1995), HIV-2 (Gao et al., 1992), EIAV (Perry et al., 1992) and CAEV (Wain-Hobson et al., 1995). As the previous sentence shows, most have been described for members of the lentiviral group and much less frequently for other retroviruses. To some extent this undoubtedly reflects the current passion with anything lentiviral, meaning that there are huge data bases for the lentiviruses with respect to other retroviruses, particularly collections of PCR amplified material. However, it probably reflects a penchant of the lentiviral reverse transcriptase (Martinez et al., 1995). Note for the time being the absence of published examples for BIV, FIV and visna virus. For the record it is not an artefact confined to viral cultures as G → A hypermutants have been identified in uncultured PCR amplified material (Gao et al., 1992; Li et al., 1991; Pelletier et al., 1995).

```
AGCGGGGGAA TGATGGAGAA AGGAGAAATG
....A..A.. .A..A..A.. .AA.A.-..A
.......A.. AA..AA.A.. .AA.A....A


GCTGGGGAAA GATGGAGAGA GGAGAAATAA
.......... A..A....A. ..........
...AA.A... A...A.A.A. AA.A......
.......... A..AA...A. .A.A......
```

Figure 1. Some examples of hypermutated HIV gp120 env sequences.

Simple Criteria for Identifying G → A hypermutants

Nucleic acid sequence:

1) Monotony of G → A transitions with respect to the viral plus strand (Pathak and Temin, 1990; Vartanian et al., 1991). Only one recorded case of the phenomenon occurring during minus strand synthesis C → T hypermutation with respect to plus strand, (Pathak and Temin, 1990; Vartanian et al., 1991). Hypermutants may be occasionally accompanied by a few (<5%) other substitutions (Martinez et al., 1994).
2) All parts of the retroviral genome are vulnerable. There is one example of a provirus hypermutated throughout all 10 kb (Borman et al., 1995; Pelletier and Wain-Hobson, unpublished data).
3) Given that purine-purine transitions are the most frequent of all substitutions for lentiviruses, overall the number of G → A transitions per sequence should be ≥5 while the transition frequency should be >5% of the number of Gs. Up to 60% of Gs may be substituted (Wain-Hobson et al., 1995).
4) The distribution of substitutions may be confined to a very small region, say 50 bp (Delassus et al., 1991). Equally, they may be distributed in an erratic manner throughout the genome (Wain-Hobson et al., 1995). Finally hypermutation may be sustained throughout minus strand synthesis resulting in approximately 30% G substitution over a 10 kb proviral sequence (Borman et al., 1995; Pelletier and Wain-Hobson, unpublished data).

5) G → A transitions are associated with dinucleotide context declining in the order GpA > GpG > GpT ≥ GpC (Vartanian et al., 1991). Occasionally a few examples have GpG>GpA (Fitzgibbon et al., 1993; Vartanian et al., 1994). More complex sequence context rules have been proposed (Borman et al., 1995). This author is not convinced the latter will hold up.
6) Hypermutants may be accompanied by small deletions of 1–5 bases. Large deletions and small insertions (1–3 bases) are rarer (Pezo et al., 1996; Vartanian et al., 1991; Vartanian et al., 1994).

Protein sequence:

7) In phase stop codons resulting from tryptophan (TGG) to stop codons (TAA, TAG and TGA); see Figure 2.
8) Unusual proportions of certain amino acid substitutions, e.g., depletion of glycine (G → R, S, D, E and sometimes N or K), arginine (R → K, and less R → H or Q), serine (S → N), aspartic and glutamic acids (D → N and E → K respectively), valine (V → I), alanine (A → T), methionine (M → I) and cysteine (C → Y); see Figure 2.

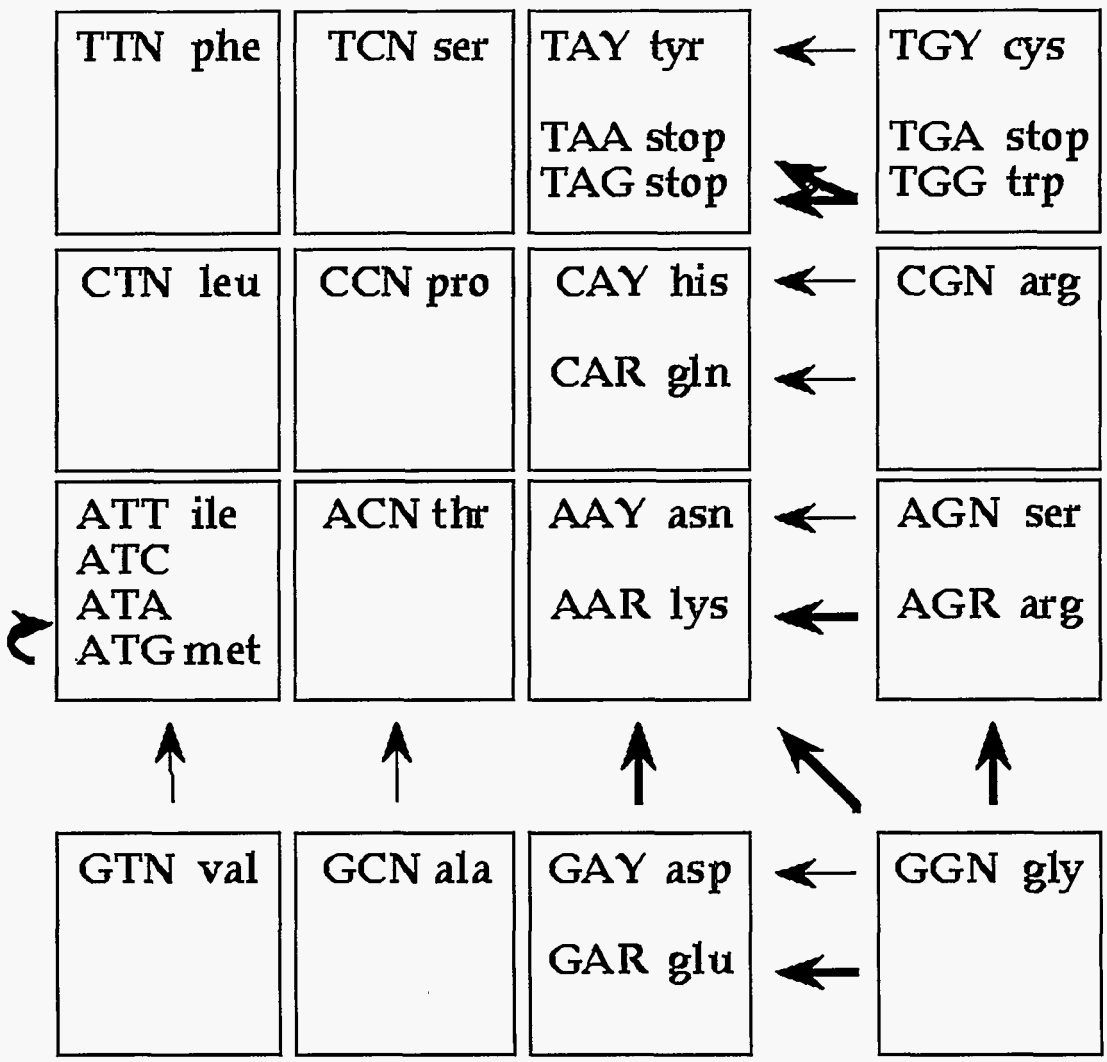


Figure 2. Amino acid substitutions resulting from G → A hypermutation. The intensity of the arrows indicates the relative frequency based generally on dinucleotide context.

Mechanism

G → A hypermutation results when minus strand DNA synthesis coincides with an increase in the intracellular [dTTP]/[dCTP] ratio.

Evidence

1) The two or so bases preceding the site of a misincorporation strongly influence the kinetics of misincorporation. Thus the strong dinucleotide context (Vartanian et al., 1991) indicates that it occurs during reverse transcription of plus stranded RNA into complementary DNA.
2) A sequence with a number of C → T transitions (with respect to the plus strand, which are but G → A transitions with respect to the minus strand) was identified along with many more G → A hypermutants (Vartanian et al., 1991). This indicated that the phenomenon could occur during both minus and plus DNA strand synthesis. The properties of RNA:DNA hybrids in vitro are more conducive to generation of rG:dT mismatches with respect to dG:dT mismatches in DNA:DNA duplexes (Sala et al., 1995).
3) G → A hypermutants may be produced in vivo by addition of deoxythymidine (dThd) to the culture supernatant (Vartanian et al., unpublished data). dThd is scavenged, and kinased up to dTTP which has a negative effect on ribonucleotide reductase reduction of CDP to dCDP. The net effect is to increase the [dTTP]/[dCTP] ratio.
4) G → A hypermutants may be simply reproduced in an in vitro reaction using RNA, purified RT and strongly biased [dTTP]/[dCTP] ratios (Martinez et al., 1994). They may also be made during endogenous strong stop synthesis (Martinez et al., 1994; Vartanian et al., unpublished data).
5) Many types of hypermutants can be produced in vitro by manipulation of the dNTP pools (Martinez et al., 1994; Vartanian et al., unpublished data). Thus the mutant spectrum is restricted by the intracellular milieu and much less the HIV-1 RT.
6) The extent of in vitro hypermutation was positively correlated to the magnitude of the [dTTP]/[dCTP] ratio (Martinez et al., 1994). The dinucleotide preference was reproduced at relatively low [dTTP]/[dCTP] ratio in the endogenous RT reaction (Vartanian et al., unpublished data).
7) HIV-1 RT is more sensitive in vitro to [dTTP]/[dCTP] imbalances than the Moloney murine leukemia virus (MoMLV) (Martinez et al., 1995). This helps explain the greater frequency of G → A hypermutants in lentiviral sequence data sets.
8) The clustering of rG:dT mismatches is partly explained by the observation that prior rG:dT mismatches increase the frequency of subsequent mismatches (Sala et al., 1995).

Additional negative arguments

1) Not associated with NTP imbalances during transcription. Given the pattern of reverse transcription the distribution of G → A transitions through contiguous U3 and R of the LTR proscribes a transcriptional origin (Vartanian et al., 1994).
2) Not related to excessive dUTP incorporation. Firstly hypermutants have been described for lentiviruses encoding a dUTPase (EIAV and CAEV, (Perry et al., 1992; Wain-Hobson et al., 1995)). Secondly for those without a dUTPase, such as HIV, the RT shows much greater discrimination against rG(template):dUTP as opposed to rGt:dTTP in vitro (Martinez et al., 1995). Finally, a dUTPase minus FIV mutant failed to generate G → A hypermutants (Lerner et al., 1995).
3) Not associated with a mutant polymerase. Hypermutants have been identified in a single cycle of reverse transcription with a "wild type" RT (Mansky, 1996; Pathak and Temin, 1990). Hypermutants can be produced in vitro using E. coli expressed RT derived from infectious molecular clones (Martinez et al., 1994).

Literature and Data Base

Overall hypermutants are rare, probably reflecting the rigours of purifying selection. There are probably fewer than 100 examples, many having been alluded to above. Some researchers relegate them

to the drawer as they are frequently defective. Others submit them to the data base without commenting on them in the conventional literature. Some examples go unperceived (Table 1).

Table 1 Some unsung or little known G → A hypermutants

Clone designation	Region of HIV genome*	Reference
ELI.03, TRA.18	env V1-V2	Goodenow et al 1989
clone 229	env V3 loop	LaRosa et al 1990
L3.14 (partially)	nef	Delassus et al 1991
YU, 1 clone	env V3-C4	Yu et al 1991
81wk:1–1, 1–3, 1–4, 1–11	SIVmac V1-V4	Overbaugh et al 1991
FOU 29.03.89 data set, 2 clones	p24 gag	Meyerhans et al 1991
FOI 21.05.90 data set, 1 clone		
MM152–12	SIVmac gag	Chen et al 1992
MM179–02, -10, -16, -22, -33		
316LSS3env	SIVmac V1-V5	Kodama et al 1993
D25/+24	pol	Najera et al 1995
A9 (Acc no. U28514)	SIVtan V3-V5	Mueller et al 1995
DH1 (partially)	nef	Huang et al 1995
BWB-11 (partially), -33	env	Monken et al 1995
CX-B (partially) patient Q23	V2 loop	Poss et al 1995
SP-2–203	nef	Michael et al 1995
D1.01	protease	Barrie et al 1996
pMCE10.86	LTR	Estable et al 1996
pMCE29.1		
HP93A1, A2, B1, B2, C1, C2	nef	Mariani et al 1996
HP95A1, A2		
HP83B1, B2 (perhaps)		

*HIV-1 unless otherwise stated. The table includes those sequences which passed into the literature with little or no comment. Other G → A hypermutants have been found for SNV (Pathak and Temin, 1990), HIV-1 (Fitzgibbon et al., 1993; Mansky, 1996; Vartanian et al., 1991; Vartanian et al., 1994), SIV (Johnson et al., 1991; Pelletier et al., 1995), HIV-2 (Gao et al., 1992), EIAV (Perry et al., 1992) and CAEV (Wain-Hobson et al., 1995). This list is probably not exhaustive, representing all those recognized by the author.

When hypermutation is not very intense or confined to a small region the molecular products cannot be assumed to be defective (Martinez et al., 1996). Thus, hypermutants should not be incorporated in phylogenetic trees, for they are generated in a single cycle of replication. There never were intermediates as is implied when constructing phylogeny. Doing so can lead to spurious conclusions or groupings (Mariani et al., 1996).

As mentioned above, one example was noted for SNV (murine C type retrovirus group) (Pathak and Temin, 1990). One case has been described for HTLV-1 being limited to 5 G → A transitions (4 GpA, 1 GpG) within a 110 bp stretch (i.e. 4.2% substitution frequency). In this context it is worth noting that during the endogenous cDNA reaction HTLV-1 hypermutants may be produced by using biased [dNTP]s (Vartanian et al., unpublished data).

And the Pararetroviruses?

This term covers the hepadnaviruses of mammals, reptiles and birds as well as the plant badnaviruses. Perhaps the most celebrated viruses of each group are human hepatitis B virus (HBV) and cauliflower mosaic virus (CaMV) respectively. Although the infectious form harbours a DNA genome, intracellular replication passes through reverse transcription of a plus strand RNA pre-genome. Are there any hypermutants? None have been published, although two such HBV genomes have been sequenced (Günther et al., unpublished data). Hypermutation was noted in the viral plus strand, while substitutions were erratically distributed throughout the genome. The overall substitution frequencies were 12% and 26%. Both genomes encoded in phase stop codons. The dinucleotide preference for transitions was GpA>GpG>GpC GpT. Such characteristics are typical of retroviral hypermutation, indicating that HBV replication is also vulnerable to intracellular [dTTP]/[dCTP] imbalances.

A → G Hypermutation

This represents an even rarer form of retroviral hypermutation, only a few examples being published to date. The initial example was identified within the U3 region of an avian leukosis virus (ALV) provirus (Felder et al., 1994), while more recent examples come from ALV (Hajjar and Linial, 1995) and SNV (Kim et al., 1996). They most certainly arose via post transcriptional deamination of adenine moieties in the RNA genome by double stranded adenosine deaminase. Such a phenomenon was first noted for measles virus genomes in cases of subacute sclerosing panencephalitis, leading to intervention in the post transcriptional modification of some cellular mRNAs. Deamination of adenine leaves inosine, which preferentially base pairs with cytosine, leading to A → G transitions. For review see Bass, 1995. Given the high A content of HIV, the extensive secondary structure in the region 1–600 and RRE, as well as the size of the HIV sequence data, it is perhaps surprising that no good example of A → G hypermutation exists. There has been discussion of an A/I change at position 27 of TAR (Blanchard et al., 1992; Klaver and Berkhout, 1994; Sharmeen et al., 1991).

Acknowledgements

Thanks to Drs. Eric Pelletier and Yoc Hoa Tran for computer analyses and Ben Berkout for email.

References

Barrie, K. A., Perez, E. E., Lamers, S. L., Farmerie, W. G., Dunn, B. M., Sleasman, J. W., and Goodenow, M. M. (1996). Natural variation in HIV-1 protease, Gag p7 and p6, and protease cleavage sites within gag/pol polyproteins: amino acid substitutions in the absence of protease inhibitors in mothers and children infected by human immunodeficiency virus type 1. *Virology* **219**, 407–416.

Bass, B. L. (1995). An I for editing. *Curr. Biol.* **5**, 598–600.

Blanchard, A. D., Powell, R., Braddock, M., Kingsman, A. J., and Kingsman, S. M. (1992). An adenosine at position 27 in the human immunodeficiency virus type 1 trans-activation response element is not critical for transcriptional activation by Tat. *J. Virol.* **66**, 6769–6772.

Borman, A. M., Quillent, C., Charneau, P., Kean, C. M., and Clavel, F. (1995). A highly defective HIV group O provirus: evidence for the role of local sequence determinants in hypermutation during negative strand DNA synthesis. *Virology* **208**, 601–609.

Chen, Z. C., Shen, L., Miller, M. D., Ghim, S. H., Hughes, A. L., and Letvin, N. L. (1992). Cytotoxic T lymphocytes do not appear to select for mutations in an immunodominant epitope of simian immunodeficiency virus gag. *J. Immunol.* **149**, 4060–4066.

Delassus, S., Cheynier, R., and Wain-Hobson, S. (1991). Evolution of human immunodeficiency virus type 1 nef and long terminal repeat sequences over 4 years in vivo and in vitro. *J. Virol.* **65**, 225–231.

Estable, M. C., Bell, B., Merzouki, A., Monyaner, J. S. G., O'Shaughnessy, M. V., and Sadowski, I. J. (1996). Human immunodeficiency virus type 1 long terminal repeat variants from 42 patients

representing all stages of infection display a wide range of polymorphism and transcription activity. *J. Virol.* **70**, 4053–4062.

Felder, M. P., Laugier, D., Yatsula, B., Dezélée, P., Calothy, G., and Marx, M. (1994). Functional and biological properties of an avian variant long terminal repeat containing multiple A to G conversions in the U3 sequence. *J. Virol.* **68**, 4759–4767.

Fitzgibbon, J. E., Mazar, S., and Dubin, D. T. (1993). A new type of G → A hypermutation affecting human immunodeficiency virus. *AIDS Res. Hum. Retroviruses* **9**, 833–838.

Gao, F., Yue, L., White, A. T., Pappas, P. G., Barchue, J., Hanson, A. P., Greene, B. M., Sharp, P. M., Shaw, G. M., and Hahn, B. H. (1992). Human infection by genetically diverse SIVsm-related HIV-2 in West Africa. *Nature* **358**, 495–499.

Goodenow, M., Huet, T., Saurin, W., Kwok, S., Sninsky, J., and Wain-Hobson, S. (1989). HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. *J. Acquir. Immune Def. Syndr.* **2**, 344–352.

Günther, S., Sommer, G., Plikat, U., Iwanska, A., Wain-Hobson, S., Will, H., and Meyerhans, A. (unpublished data).

Hajjar, A. M., and Linial, M. L. (1995). Modification of retroviral RNA by double-stranded RNA adenosine deaminase. *J. Virol.* **69**, 5878–5882.

Huang, Y., Zhang, L., and Ho, D. D. (1995). Characterization of nef sequences from long-term survivors of human immunodeficiency virus type 1 infection. *J. Virol.* **69**, 93–100.

Johnson, P. R., Hamm, T. E., Goldstein, S., Kitov, S., and Hirsch, V. M. (1991). The genetic fate of moleculary cloned simian immunodeficiency virus in experimentally infected macaques. *Virology* **185**, 217–228.

Kim, T., Mudry Jr., R. A., Rexrode II, C. A., and Pathak, V. K. (1996). Retroviral mutation rates and A-to-G hupermutations during different stages of retroviral replication. *J. Virol.* **70**, 7594–7602.

Klaver, B., and Berkhout, B. (1994). Premature strand transfer by the HIV-1 reverse transcriptase during strong-stop DNA synthesis. *Nucleic Acids Res.* **22**, 137–144.

Kodama, T., Mori, K., Kawahara, T., Ringler, D. J., and Desrosiers, R. C. (1993). Analysis of simian immunodeficiency virus sequence variation in tissues of rhesus macaques with simian AIDS. *J. Virol.* 67, 6522–6534.

LaRosa, G. J., Davide, J. P., Weinhold, K., Waterbury, J. A., Profy, A. T., Lewis, J. A., Langlois, A. J., Dreesman, G. R., Boswell, R. N., Shadduck, P., Holley, L. H., Karplus, M., Bolognesi, D. P., Matthews, T. J., Emini, E., and Putney, S. D. (1990). Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. *Science* **249**, 932–935.

Lerner, D. L., Wagaman, P. C., Phillips, T. R., Prospero-Garcia, O., Henriksen, S. J., Fox, H. S., Bloom, F. E., and Elder, J. H. (1995). Increased mutation frequency of feline immunodeficiency virus lacking functional deoxyuridine-triphosphatase. *Proc. Natl. Acad. Sci. USA.* **92**, 7480–7484.

Li, Y., Kappes, J. C., Conway, J. A., Price, R. W., Shaw, G. M., and Hahn, B. H. (1991). Molecular characterization of human immunodeficiency virus type 1 cloned directly from unclutured human brain tissue: identification of replication-competent and -defective viral genomes. *J. Virol.* **65**, 3973–3985.

Mansky, L. M. (1996). The mutation rate of human immunodeficiency virus type 1 is influenced by the vpr gene. *Virology* **222**, 391–400.

Mariani, R., Kirchhoff, F., Greenough, T. C., Sullivan, J. L., Desrosiers, R. C., and Skowronski, J. (1996). High frequency of defective nef alleles in a long-term survivor with non-progressive human immunodeficiency virus type 1 infection. *J. Virol.* **70**, 7752–7764.

Martinez, M. A., Pezo, V., Marlière, P., and Wain-Hobson, S. (1996). Exploring the functional robustness of an enzyme by in vitro protein evolution. *EMBO J.* **15**, 1203–1210.

Martinez, M. A., Sala, M., Vartanian, J. P., and Wain-Hobson, S. (1995). Reverse transcriptase and substrate dependence of the RNA hypermutagenesis reaction. *Nucleic Acids Res.* **14**, 2573–2578.

Martinez, M. A., Vartanian, J. P., and Wain-Hobson, S. (1994). Hypermutagenesis of RNA using human immunodeficiency virus type 1 reverse transcriptase and biased dNTP concentrations. *Proc. Natl. Acad. Sci. USA*. **91**, 11787–11791.

Meyerhans, A., Dadaglio, G., Vartanian, J. P., Langlade-Demoyen, P., Frank, R., Asjo, B., Plata, F., and Wain-Hobson, S. (1991). In vivo persistence of a HIV-1-encoded HLA-B27-restricted cytotoxic T lymphocyte epitope despite specific in vitro reactivity. *Eur. J. Immunol.* **21**, 2637–2640.

Michael, N. L., Chang, G., D'Arcy, L. A., Tseng, C. J., Birx, D. L., and Sheppard, H. W. (1995). Functional characterization of human immunodeficiency virus type 1 nef genes in patients with divergent rates of disease progression. *J. Virol.* **69**, 6758–6769.

Monken, C. E., Wu, B., and Srinivasan, A. (1995). High resolution analysis of HIV-1 quasispecies in the brain. *AIDS*. **9**, 345–349.

Mueller-Trutwin, M. C., Corbet, S., Dias-Tavares, M., Herve, V. M. A., Nerrienet, E., Georges-Courbot, M. C., Saurin, W., Sonigo, P., and Barré-Sinoussi, F. (1995). HIV sequence data base.

Najera, I., Holguin, A., Quinones-Mateu, M. E., Munoz-Fernandez, M. A., Najera, R., Lopez-Galindez, C., and Domingo, E. (1995). pol gene quasispecies of human immunodeficiency virus: mutations associated with drug resistance in virus from patients undergoing no drug therapy. *J. Virol.* **69**, 23–91.

Overbaugh, J., Rudensey, L. M., Papenhausen, M. D., Benveniste, R., and Morton, W. R. (1991). Variation in simian immunodeficiency virus env is confined to V1 and V4 during progression to simian AIDS. *J. Virol.* **35**, 7025–7031.

Pathak, V. K., and Temin, H. M. (1990). Broad spectrum of in vitro forward mutations, hypermutations, and mutational hotspots in a retrovital shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. *Proc. Natl. Acad. Sci. USA*. **87**, 6019–6023.

Pelletier, E., Saurin, W., Cheynier, R., Letvin, N. L., and Wain-Hobson, S. (1995). The tempo and mode of SIV quasispecies development in vivo calls for massive viral replication and clearance. *Virology* **208**, 644–652.

Pelletier, E., and Wain-Hobson, S. (unpublished data).

Perry, S. T., Flaherty, M. T., Kelley, M. J., Clabough, D. L., Tronick, S. R., Coggins, L., Whetter, L., Lengel, C. R., and Fuller, F. (1992). The surface envelope protein gene region of equine infectious anemia virus is not an important determinant of tropism in vitro. *J. Virol.* **66**, 4085–4097.

Pezo, V., Martinez, M. A., and Wain-Hobson, S. (1996). Fate of direct and inverted repeats in the RNA hypermutagenesis reaction. *Nucleic Acids Res.* **24**, 253–256.

Poss, M., Martin, H. L., Kreiss, J. K., Granville, L., Chohan, B., Nyange, P., Mandaliya, K., and Overbaugh, J. (1995). Diversity in Virus populations from genital secretions and peripheral blood from women recently infected with human immunodeficiency virus type 1. *J. Virol.* **69**, 8118–8122.

Sala, M., Wain-Hobson, S., and Schaeffer, F. (1995). HIV-1 reverse transcriptase tG:T mispair formation on RNA and DNA templates with mismatched primers: a kinetic and thermodynamic study. *EMBO J.* **14**, 4622–4627.

Sharmeen, L., Bass, B., Weintraub, H., and Groudine, M. (1991). Tat-dependent adenosine-to-inosine modification of wild-type transcativation response RNA. *Proc. Natl. Acad. Sci. USA*. **88**, 8096–8100.

Vartanian, J. P., Meyerhans, A., Åsjö, B., and Wain-Hobson, S. (1991). Selection, recombination and G → A hypermutation of human immunodeficiency virus type 1 genomes. *J. Virol.* **65**, 1779–1788.

Vartanian, J. P., Meyerhans, A., Sala, M., and Wain-Hobson, S. (1994). G → A hypermutation of the HIV-1 genome: evidence for dCTP pool imbalance during reverse transcription. *Proc. Natl. Acad. Sci. USA*. **91**, 3092–3096.

Vartanian, J. P., Plikat, U., Henry, M., Mahieux, R., Guillemot, L., Barré-Sinoussi, F., Meyerhans, A., and Wain-Hobson, S. (unpublished data).

Wain-Hobson, S., Sonigo, P., Guyader, M., Gazit, A., and Henry, M. (1995). Erratic G → A hypermutation within a complete caprine arthritis-encephalitis virus (CAEV) provirus. *Virology* **209**, 297–303.

HIV Alignments, Database Searches, and Structure Predictions

Gerald Myers and Andrew Farmer

MS K710, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Analyses of more distantly related HIV and SIV sequences that take as their point of departure an alignment, either of the nucleic acid or amino acid sequences, will only be as sound as the alignment, which is itself an hypothesis. For this reason, many sequence analyses are conducted over only "unambiguously alignable" stretches of sequence, typically stretches for which the similarities are 50% or greater, and the information in the more varied regions (similarities less than 30%) is lost to the analysis. This is unfortunate insofar as structural information is typically derived from distantly related, not closely related, sequences. Given the diversity of primate immunodeficiency viral (PIV) sequences, alignments in the database publications up to this year were constrained of necessity to highly related subgroups of viral types, which were then "apposed" by eye. These safe, but restricted, alignments could not support some analyses based upon the entire set of PIV sequences. This year, in Parts I, II and III, we have brought into play a new alignment strategy that holds some promise for simultaneously and objectively aligning all members of the primate immunodeficiency virus family.

An additional reason for exploring new approaches to HIV and SIV sequence alignment concerns computational time: multiple alignment, viewed as an extension of pairwise alignment, requires time proportional to the average length of the sequences raised to power k, where k is the number of sequences being aligned. In the approach described below, sequences are aligned in time linearly proportional to k, the number of sequences. There are further advantages to the method. Most alignment strategies are "progressive", which is to say that the alignment unfolds from the pairs of most similar sequences to the pairs of most dissimilar sequences; the essence of this approach is captured by Doolittle's dictum—"once a gap always a gap" [1]. McClure and coworkers critique twelve different alignment methods, most of which are progressive, according to their abilities to correctly identify ordered series of motifs in highly divergent proteins that have been experimentally studied [2]. They find that no single approach is superior to all others, and most are time-consuming. Some of the newer multiple alignment programs are intentionally not progressive, partly for the reason that progressive alignments may be trapped by local optima, partly because phylogenetic inferences are implicitly assumed. The Hidden Markov Method (HMM) approach, which we utilize and describe herein, is not progressive; instead, it emphasizes position-specific probability distributions of character states, hence a gap in one portion of the alignment may be scored differently than a gap in another portion of the aligment. Most alignment programs have position-independent scoring schemes, which are unrealistic in the case of most proteins since they are composed of both conserved regions and indel-rich variable regions. As HMM is centered upon the columns of information, this approach, referred to as a "generalized profile", is indifferent to the relatedness of pairs of sequences [3,4]. On the other hand, because a probability distribution over the 20 amino acids must be constructed at each position in the sequence, HMM is employed with large sets (40 or more taxa) of highly divergent sequences such as is seen with HIVs and SIVs. Furthermore, an assumption of independence between sites is made. The latter assumption is not universally defensible, however neither is it new to sequence analyses.

At this point in the discussion, it will be helpful to briefly recall what a Markov chain is and in what sense the sequence alignment problem is said to be "hidden." A first order Markov process is one in which the state at time t is a probabilistic function of the state at time $t-1$. We *think* that this is a reasonable assumption for viral sequence evolution. Many phenomena describable by a Markov process are observable, *e.g.*, changes in weather where a complete history of weather is on record [5]; when the changes of state are themselves not observable, the process is said to be hidden. In the case of HIV sequences, we do not possess a history of the intermediate states through which sequences have evolved. Fitch's notion of a *covarion*, the set of concomitantly variable codons, can illustrate the hiddeness: we imagine that evolutionary changes shift the makeup of the covarion, at

one time a position be invariant, at another time being variable; one state representing one covarion, another state representing another covarion. These states are not observed. The overall problem is to formally arrive at a probabilitistic model that will satisfactorily account for the sequences that are known. Specific aspects of the problem, in addition to facilitating alignments, are to derive a suitable "consensus" sequence, to classify a sequence (is it distantly related or totally unrelated?), to support structural analysis, and to gain insight into the "hidden" evolutionary process.

Hence the HMM approach leads to a model for the sequence set that has been analyzed. An example of an HMM-generated model, or so-called architecture, is shown on the cover of this year's compendium. With subsequent database searches, this model—in the form of a "most likely" sequence, or *discriminator*—embodies all of the information contained in the data set, not merely one particular sequence. As we shall see, this consensus-like sequence is useful for database searching (below and accompanying Part IV section concerned with Molecular Mimicry). We have also coupled the HMM-generated model to prediction analyses of protein structure using an array of contemporary algorithms. Eventually, the sequence alignment and the structure prediction will become intertwined in an effort to optimize the alignment.

Be forewarned that the method is still in its infancy and that our utilization is at the most elementary level: refinement of the approach, especially to extend the analysis to nonprimate lentiviruses and other retroviruses, entails extensive parameterization studies such as those being undertaken by McClure and coworkers [6]. Issues such as optimum model length, size of training set, etc. are not taken up in this analysis, which was restricted to just primate lentiviral sequences.

In the following text, we first describe in some detail the HMM approach to alignment as we have applied it in this compendium, especially in Parts I and II, then we turn to discussions of database searching and protein structure prediction.

A. MULTIPLE SEQUENCE ALIGNMENT OF HIVS and SIVS USING HMM

The Hidden Markov Model (HMM), as it has been applied to sequence analysis, has many similarities to what is called a "profile" [7,8] in terms of the information that it captures concerning a set of related sequences. In a sense, each can be thought of as an extended consensus sequence in which the information retained at each position includes the frequency with which each possible base or amino acid residue is seen in the sequence set at that position. The HMM is constructed from a number of successive nodes generally corresponding to the columns of positional homology of an alignment; each of these nodes contains a match state, an insert state and a delete state (see figure on the cover of this compendium). Match states correspond to simple amino acid (or base) substitutions; insert and delete states are self-explanatory. Associated with each of the states in the model are vectors of probabilities that specify the likelihood with which the system might pass to each member of the set of next possible states; these are referred to as transition probabilities. Also associated with match states and insert states are vectors of probability specifying the likelihood that the system will generate or "emit" each possible amino acid or nucleotide when in that state; delete states allow for the possibility that a sequence not have a character in a certain column. Altogether, there will be three probability matrices to describe the model—the transition matrix, the emission matrix, and the initial state matrix.

The resultant architecture of the HMM allows one to establish a correspondence between the characters of a given sequence and the states of the model. The succession of the characters in the sequence will thus determine a path through the states of the model, and associated with this path will be a likelihood determined both by the probabilities of transition between successive states and the probability that each state has for generating the character that has been assigned to it. Provided that all the probabilities in the model, including both transition and emission probabilities, are non-zero, then each path through the model that is permissible according to the rules governing transitions from one node of the model to the next will have a non-zero probability of generating the given sequence. The task of finding the optimal path through the model for a given sequence, i.e. the path with the highest likelihood, can be thought of as aligning the sequence to the model, and may be solved using dynamic programming techniques.

The most important differences between the profile and the HMM lie not in the resultant information structures, but in the means by which these structures are generated from the sequence data. As with an ordinary consensus sequence, the profile is generated from a set of sequences whose alignment has been determined by some independent means. The parameters for describing an HMM can also be derived from a given alignment in this manner. More importantly, however, there exists an algorithm for HMMs that allows one to determine the parameters of the model having the highest likelihood (at least within the neighborhood of the initial model) given a set of unaligned sequences. This approach is quite similar to certain techniques used in connection with artificial intelligence applications, and is known as "training" the model. The general procedure for training makes use of an Expectation-Maximization algorithm, of which there are several [3,4,6].

Speaking generally, the algorithm used in training the parameters of an HMM involves an iterative approach that uses an initial model to estimate an alignment of the given set of sequences, then uses this alignment to re-estimate the model, and so on until the estimates converge to an optimum. For example, if we are given a set of protein sequences that are thought to be related, a good estimate for an initial model can be made by using the frequency distribution of amino acids in the unaligned set as a vector of probabilities assigned to all the match states and insert states of the model; transition probabilities between the states of the initial model can be assigned arbitrarily, or using a prior assessment of the relative frequencies of indel events. All of the sequences in the given set will now be aligned in turn to the model, finding the path through the model that maximizes the likelihood for the given sequence; by aligning all the sequences in the set to the model in this pairwise fashion, one transitively defines a multiple sequence alignment of the sequences to one another. The multiple sequence alignment thus created can be used for an estimate of the parameters of the HMM, by counting the frequency of occurrence of each amino acid at each position of the alignment and the frequency of indel events across the alignment. This adjusted HMM then serves as a model for another round of alignment, and so on. It can be shown that this process is guaranteed to converge to a local maximum of the likelihood function.

To address the problem of guaranteeing convergence to a global maximum for this function, a variation of the simulated annealing algorithm can be applied at each step of the iterative algorithm; this basically allows a stochastically generated sub-optimal alignment to be chosen for the re-estimation of the model's parameters, where the sub-optimality of the alignment decreases to zero with successive iterations of the re-estimation procedure.

As should be clear from the preceding discussion, the model can be used to generate a multiple sequence alignment of sequences, including sequences not belonging to the set used to train the parameters of the model. A distinct advantage to using the HMM over the standard dynamic programming algorithm for multiple sequence alignments is that since one is really performing a set of pairwise comparisons of the sequences to the model, the time and memory requirements increase only linearly with the number of sequences, as opposed to exponentially with dynamic programming. It follows that it is relatively easy to add a new sequence to the alignment and rebuild the model; experimentally-derived information can also be added to the model (known as *priors*) with relative ease.

A good general introduction to the basic ideas of HMMs (not oriented, however, toward sequence analysis) is reference 5 below. We have employed the HMMER implementation that is publicly available [9–10](*eddy@genetics.wustl.edu*). Another HMM suite that can be obtained is SAM (*http://www.cse.ucsc.edu/research/compbio/sam.html*). The SAM Web site contains a number of links to papers concerned with HMMs and sequence analysis. These programs were originally applied to highly studied data sets, "validation" sets [3,4,6] (globins, EF-hand proteins, kinases, and proteases), for which extensive experimentally-based data were available to help assess the alignment results. With HIV and SIV sequences, the results of the approach must be critiqued by scutinizing motifs—do cysteines, cleavage sites, and potential glycosylation sites in envelope align, for example? This can be problematic as it is not preordained, for instance, that all sequons need align [11]. Another approach to critiquing the HMM-generated alignment involves construction of blocks using representative PIV sequences (described below). Finally, it is necessary to compare scores from matching individual HIV

and SIV sequences to the "most likely", or *discriminator*, sequence, the HMM equivalent of a consensus: as we shall report in the section to follow, our model has been "overfitted" to HIV-1 sequences; nevertheless, the boundaries and motifs appear to be satisfactorily aligned.

Approximately 400 HIV and SIV sequences were trained using HMMER version 1.8. The frequency distribution of amino acids in the unaligned HIV and SIV sequences was used as input in place of the default distribution derived from the PIR database. We should first consider the success or failure of the approach with respect to identifying motifs. The envelope protein alignment of an HIV2 sequence (SBL/ISY) to the HMM consensus is instructive in this regard (figure 1): 1) cysteines, denoted by '*', and potential N-linked glycosylation sites, denoted by '^', are aligned as we might expect them to be aligned when doing ordinary alignment assisted by manual input; 2) four noncontiguous residues found in certain HIV-1 subtype B sequences to be essential for CD4 interaction, D, E, W and D (see pp. II-1,2 of NOV 95), are aligned in almost all HIVs and SIVs (tryptophans are in general highly conserved); 3) the gp120/gp41 cleavage site is aligned, as we expect, although an alternative cleavage site may be used in some HIV-2s.

Further confirmation of the alignment comes from BLOCKS analysis: using representative sequences from the Part II HMMER-generated alignments, blocks—gapless arrays of multiply aligned conserved sequences—were constructed using the BLOCKMAKER program [12,13] (http://www.blocks.fhcrc.org). Two different programs are employed by BLOCKMAKER, the MOTIF program and the GIBBS Sampler program. Boundaries for the created blocks were highly conserved in the HMM alignments (Part II), however blocks based on envelope sequences largely coincide with conserved domains, hence other motifs (cysteines, glycosylation sites) must provide the main support for alignment over the more varied regions.

A difficulty encountered by the HMM (and every) alignment method is large indels; we have the least confidence in those. To the extent that these stretches may have arisen through acquisition of genetic material, they may not be intrinsically alignable as they may not be homologous. The reader may want to compare the V1-V2 region alignment in figure 1 with one manually created by Lamers et al. [14]. Alignments in previous publications may be "safer", but they are also more constrained and less informative because they were executed over just highly related sequences. Since the different alignments have different applications, both are made available on the Web site (*http://hiv-web.lanl.gov*).

Nucleotide sequence alignments, by this approach, were produced from the HMM-generated amino acid sequence alignments. The nucleotide alignment was then subjected to HMMER and a nucleotide-specific model was obtained. This approach follows in a general way the approach taken in earlier database publications that were based on the PIMA algorithm of Smith and Smith.

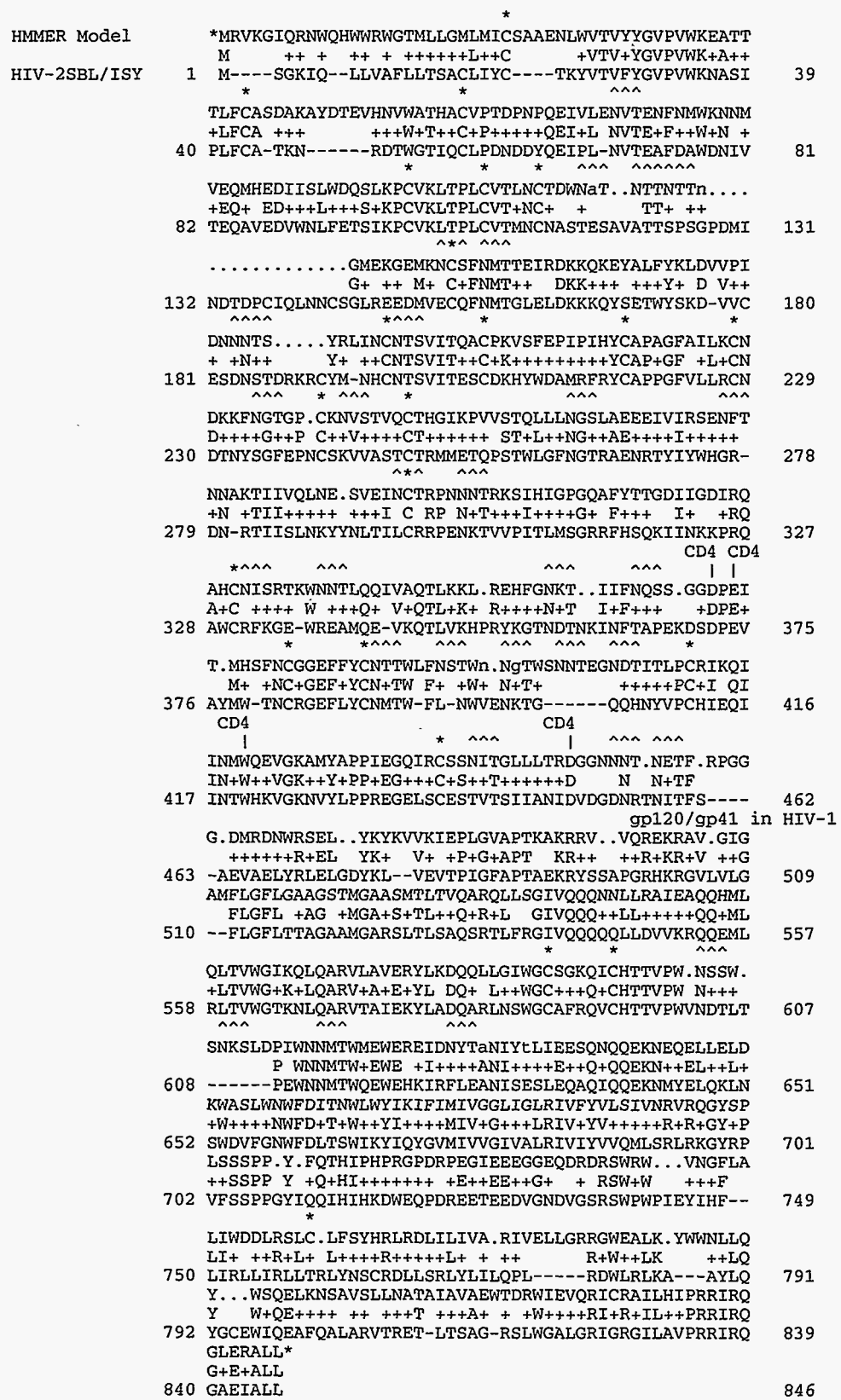
B. DATABASE SEARCHING USING AN HMM APPROACH

An important application of the HMM-generated model is in the discrimination of related sequences from non-related sequences. This is especially useful in connection with database searching. Associated with each sequence in the database is a probability, a log-odds score analogous to a BLAST score, with which the sequence could be generated by the given model. With the HMMER algorithm [10],

$$\text{score} = \log_2 \frac{P(S_i|M)}{P(S_i|R)}$$

where the alignment of each sequence in the database, S_i, is compared to both the HMM generated model, M, and a random model, R. The latter should have the same amino acid composition as the database at large and it should be as likely, *a priori*, as M. The log-odds score corrects for sequence length. A score greater than zero has a better than even chance of being significant, however, as a rule of thumb, a score must be about 20 or more to be deemed significant [9]. The distribution of likelihood scores for all the sequences in the database will provide a measure of discrimination between similar and non-similar sequences. Using the HMM for database searching has the advantage of utilising a great deal more of the information available for a family of sequences than can be captured by query techniques that force one to use only one sequence from the family or, at best, a standard consensus sequence as a query against the database.

Figure 1. Comparison of an HIV-2 Env Sequence to the HMM Model



The HMMER program, used herein, employs the Smith-Waterman (S-W) algorithm for optimizing local alignments. As with BLAST, both identities and equivalencies (conserved amino acid replacements) are scored; unlike BLAST, which is a heuristic search program, the S-W algorithm reports only the best match between the HMM consensus and a given database entry. A general discussion of these somewhat different approaches to database searching is found in reference [15]. Compositional bias may be present in the outcome of a database search, which is to say that spurious matches may arise by virtue of similar compositions only.

The comparison shown in figure 1 of a particular HIV-2 sequence to the HMM model consensus was taken from an S-W/HMMER search of the protein database using the HMM model for envelope. The score (corrected for the model length and the length of the target sequence) was highly significant, 1676, however HIV-1 scores uniformly hovered between 2700 and 2800, which tells us that the model was overfitted to HIV-1s. Scores among HIV-1s of different genetic subtypes did not significantly differ, which was satisfying, but HIV-2 and SIV scores were uniformly lower. The *maximum discrimination* option of HMMER, in contrast to the default *maximum likelihood* improves this situation somewhat [9–10]. Note in figure 1 that the reduction in score for the HIV-2 is mostly due to amino acid differences regarded to be conservative (indicated with a'+'). Although the probability distributions at the various homologous sites did not have adequate representation from the HIV-2 sequences, the alignment, as such, is reasonable. The scores over just the envelope gp41 were closer; and the scores for gag protein were much closer, 1759 for an HIV-1, 1478 for an HIV-2, showing that the fit is more inclusive.

Parallel S-W/HMMER envelope searches were conducted using a database in which HIV and SIV sequences were filtered out. The only non-zero matches to nonprimate lentiviral sequences involved Visna and its close relatives, OMVVSA and CAEV. The log-odds scores for these were less than 10; for example, from the Env gp41,

```
HMM Consensus   NNMTWMEWEREIDNYTaNIYtLIEES
                +N TW++WERE   Y +N + L+ ES
CAEV Env        DNCTWQQWERELQGYDGNLTMLLRES
```

The score for this match was 5.55, suggesting that the cutoff value of 20 should not be too rigidly applied. The highest match score in this search, 12.15, belongs to a horse skeletal muscle sodium channel alpha-subunit, which illustrates the possibility of compositional bias:

```
HMM Consensus   NTTWLFNSTWn.NgTW.SNNTEG.ND
                NTTW  N TW+ N+TW SN+T++ ND
query           NTTWYGNDTWYSNDTWNSNDTWSSND
```

In summary, the HMM-based search is fairly stringent as no significant matches were found to the HIV/SIV envelope model that were not proteins from primate immunodeficiency viruses. Selected examples of envelope matches with weak scores (less than 10) can be found in the accompanying section of Part IV concerned with molecular mimicry. HMM-generated models and database searches were also conducted for Gag, Tat, Vpr-Vpx, Vpu, and Nef, all of which are more evenly fitted than Env to all primate immunodeficiency viruses. Some brief comments regarding the search results follow:

Gag: Significant matches to nonprimate lentiviruses were more common with the HMM consensus for Gag than for Env: Visna, CAEV, EIAV, Jembrana, BIV and FIV all displayed matches, with FIV having the highest score (approximately 85). Most of these matches included the Gag zinc-finger motif, and as a result many matches above a score of 20 were also observed for proteins other than Gag. The tetrahymena cnjB gene product, for instance, scored 35.33, which was comparable to scores for some of the nonprimate lentiviral Gags. Many scores less than 20 were encountered for the zinc-finger motif, for example:

```
HMM Consensus           RKIIKCFNCGKEGHIARNCRAPRKKGCWKCGKEGH
                        R + KCFNC  EGH    C+ P  +GC  CG  GH
C. elegans RNA helicase RGPMKCFNCKGEGHRSAECPEP-PRGCFNCGEQGH
```

The log-odds score for this match was 4.67. We conclude from this search that the HMM model was especially good for picking up zinc-finger motifs of a certain kind.The so-called major homology region (MHR), which in the HMM consensus appears as IRQGPKEPFRDYVDRFYKTL, showed up only in matches with lentiviral Gags or with retroviral type D Gag sequences, such as those of SRV.

Tat: Of the nonprimate lentiviruses, only BIV and the Jembrana disease virus possessed significant match scores to the HMM consensus, 43.75 and 37.94 (versus an HIV-1 or HIV-2 score of about 280). These matches were across the second and third domains of the first coding exon, which encompass cysteine residues involved in intramolecular bonding and the R/KKGLGI motif that is thought to constitute the minimal Tat. EIAV displayed a weaker match, 10.43, and only in the second domain. These findings corroborate earlier judgments in the field that only BIV, and possibly EIAV, among nonprimate lentiviruses, possess a "true Tat" (the Jembrana virus has been sequenced subsequent to that conclusion.) Hence, FIV's match in the fourth domain, with a score of merely 5.26, is not considered Tat-specific:

```
HMM Consensus    KKRRQRRRTPQKS
                 KK RQRRR ++K+
            FIV  KKKRQRRRRKKKA
```

Vpr-Vpx: Given the paralogous relationship between Vpr and Vpx, one HMM Consensus was generated for the two proteins. The highest score against this consensus for sequences other than primate lentiviral sequences was with SA-OMVV, the Visna relative:

```
HMM Consensus    MEQAPWEfPRERIDQGWEWDPQRE
                 ME+A    PR    +G     +RE
      SA-OMVVSA  MEEA----PRR--RPG----GSRE
```

The score was 10.44. A nearly identical score was attained by a ligand for Fas antigen, but clearly due to mere compositional homogeneity:

```
HMM Consensus    GPGGWRRGPPPRNPPSRSMH
                 GPG+ RR PPP++PP+ S +
ligand for Fas   GPGQ-RR-PPPPPPPP-SPL
```

Vpu: In contrast to other HMM consensus sequences, the discriminator for Vpu was generated solely from HIV-1s. Match scores varied significantly—251 for the BAL strain but merely 45.38 for ANT70—suggesting that the model was overfitted to M group viruses. The subtype D virus ELI had a score of 229. Among non-HIV-1s, the highest score, 10.93, was found for a toxin receptor:

```
HMM Consensus    MQPLQILAIVALVVAaIIAIVVW
                 +  L+I A V LV ++  A VVW
C5a anaphylatoxin receptor  ILALVIFAVVFLVGVLGNALVVW
```

Rev sequences from SIV displayed weak similarity with the N-terminus of the HMM model for Vpu:

```
HMM Consensus    MQPLQILAI
                 +Q LQ LAI
      SIV Rev    IQQLQRLAI
```

Nef: With exception of primate immunodeficiency viral sequences, no database sequence displayed striking similarity to regions of the NEF HMM consensus. When mediocre scores were seen, they invariably resulted from compositional homogenity—cysteines, acidic amino acid residues, etc. This result is in contrast to the widespread claims regarding Nef "homologs" (i.e., similarities) in the literature of HIV.

From these preliminary studies, we conclude that the HMM approach to database searching is sensitive and specific. As a supplement to searches based on an HMM consensus, users should consider searching with COBBLER sequences deduced from the BLOCKMAKER program (http://www.blocks.fhcrc.org). Future improvements to the approach will include the introduction of structural information. We shall now turn to structure-prediction and the interplay between primary sequence alignment and structure-based alignment.

C. PROTEIN STRUCTURE PREDICTION

A promising starting point for predicting a structure for a given amino acid sequence is to determine whether that sequence is sufficiently similar to any other sequence for which biophysical data, ideally X-ray crystallographic data, is available. For sure, sequences that are 50% are more similar will have similar structures, while less similar sequences can have similar folds over core regions. The focus herein will be upon weakly similar sequences, in particular upon those of envelopes for which comparatively little biophysical data, beyond limited NMR, is available.

The earliest structure prediction algorithms, such as the Chou-Fasman algorithm, possess a predictive accuracy of no better than about 55%, partly due to the small set of known structures upon which they depend and partly due to their assumptions. Three-state predictions—helix (*H*), sheet (*E*) and coil (*C*)—are more accurate than four-state predictions that include turns (*T*); the accuracy is poorest at the ends of polypeptides and best in the core regions. Secondary stucture prediction in general is most reliable for transmembrane helices. With the buildup of the protein database and the development of more powerful algorithms, which especially take into account multiple sequence alignments, the predictive accuracy for secondary structure can now reach slightly better than 70%.

SOPM (self-optimised prediction method) is an example of a recent approach to protein secondary structure prediction [16–17]. When applied to 239 dissimilar proteins of known structure, this algorithm yields three-state prediction accuracies of 69% to 73%. On the other hand, because it involves sizeable subdatabases of sequences and their known structures, it will take longer to run than the older, less accurate algorithms. The basic ideas used in the SOPM are as follows.

First, a sliding window of a fixed size is applied to the protein sequence of unknown secondary structure to define a set of overlapping peptides. For example, suppose we are given the sequence KPQRNSKSTAAL ... with a window whose size is eight amino acids long and which is moved one amino acid over at each step. The resultant set of octapeptides will be KPQRNSKS, PQRNSKST, QRNSKSTA, RNSKSTAA, NSKSTAAL.... Note that most of the amino acids of the original sequence will belong to successive octapeptides, each differing from the previous peptide by the removal of an amino acid from one end and the addition of an amino acid to the other.

Next, each of the peptides thus derived from the query sequence is now compared to a database of peptides that has been created by similar means from a database of proteins of known secondary structure. If the peptide from the query sequence matches a peptide from the database above a certain threshold of similarity, then the similarity score is added to the conformational scores for each of the amino acids in the peptide. In our example, suppose that the first peptide KPQRNSKS matches a peptide in the database RPQRDTKS whose known structure is *HHCCCEEE*, and that the similarity score between these two peptides is 30. If this score is above the threshold parameter, then 30 will be added to the first two amino acids' helical conformational scores, to the next three amino acids' coil conformational scores and to the last three amino acids' sheet conformational scores. There may be other peptides in the database matching the query peptide with alternative predictions for the secondary structure of each of the amino acids, and all these predictive scores will be added together in each of the conformational categories, resulting in a distribution of scores over the possible secondary structure conformations. After the first query peptide has been compared, the process will continue for each of the remaining peptides in the query set. The final scores for an amino acid belonging to eight successive query peptides will thus include the scores for the comparisons of all eight of these peptides against the entire database of peptides of known structure.

After all comparisons have been made, each amino acid in the original protein will have values associated with its propensity to adopt a conformation in each of the secondary structure classes. From

the method of calculation detailed above it is clear that the empirical evidence for the prediction of the secondary structure of the amino acid weighs most heavily for that class with the highest score.

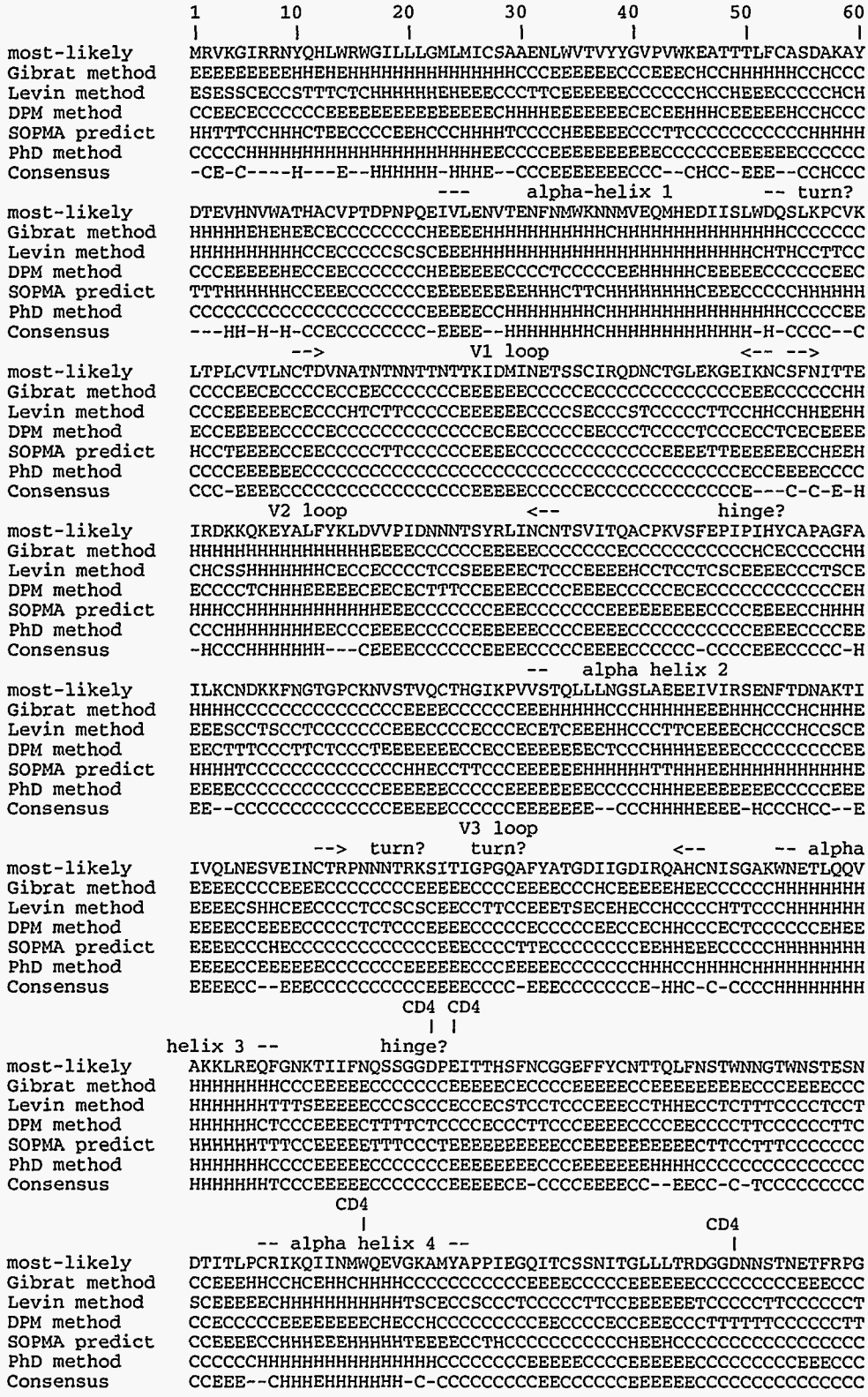
There are two additional statistics concerning the distribution of the scores over all the classes that can be revealing of the predictive power of this approach. The first is the actual magnitude of the scores for any given amino acid. If these are small relative to the cumulative scores for other amino acids, it may indicate a lack of information for the prediction of the secondary structure conformation of that amino acid. This could happen for two reasons: first, if the amino acid is within the window size to either terminus of the original protein, it will belong to proportionately fewer query peptides and have fewer comparisons with the database that could add to its score; second, the amino acid could belong to a series of peptides that for some reason are poorly represented in the database of known structures, and could thus have few comparisons to the database having a large enough similarity score to be added to the conformational scores for the amino acid. In either case, values that are low in magnitude indicate a lack of information in the database for the amino acid in its given environment.

The second statistic that is pertinent to the predictive value of a set of scores for a given amino acid is the difference between the scores of the highest and next-highest scoring classes of secondary structure. If this difference is small, it may be inferred that the information in the database for the amino acid in this particular environment is conflicting. For example, suppose that approximately half of the peptides contributing to a given amino acid's conformational scores support a helical structure, while the other half support its being classed as an element of a beta sheet. In this scenario, it is likely that the cumulative scores for helix and sheet for this amino acid would be nearly equal, and thus the difference between them would be near zero. In order to make this statistic independent of the magnitudes of the scores (which were accounted for in the former statistic), one may normalize the values by dividing the difference between the highest and next-highest scores by the magnitude of the highest score.

It is the widespread wisdom at this time to evaluate sequences, whenever possible, by more than one algorithmic approach; some methods are better for predicting helices, others for predicting sheets, etc. The SOPMA server (*http://www.ibcp.fr/predict.html*), therefore, submits a sequence to alternative methods of structure prediction—Gibrat, Levin, DPM and the PhD [18–21]—and also generates a consensus over those and the SOPM prediction itself. The HMM-generated "most likely" sequence was submitted to the SOPMA suite, producing predictions using the five individual algorithms as well as a consensus prediction, with the following results(H = helix, E = beta sheet, T = turn, and C = coil; blocks, defined by the MOTIF program in BLOCKMAKER, are indicated).

Although the HMM model has been overfitted in this case to HIV-1, the structures should be somewhat conserved across primate lentiviral boundaries. Gallaher and coworkers have explored this for the surface and transmembrane components of the lentiviral envelope [22,23]. Because the HMM-generated "most likely" sequence embodies information from hundreds of primate immunodeficiency viruses, it offers a reasonable test of their "eclectic" models derived from representative lentiviral sequences [22,23].

The Gallaher model for the surface protein (dependent primarily upon the Chou-Fasman algorithm) identifies five helical regions, all five of which are strongly or moderately predicted by the SOPMA suite: the 1st overlaps the 1st definable block in gp120 (figure 2 and Part II); the 4th is included in the 4th block and the 5th is included in the 5th block. Helices 1, 3 and 5 are strongly evident in HIV-2 sequences such as ROD. Furthermore, SOPMA suggests a small stretch of helix at the C-terminal end of the V3 loop and also at the C-terminal end of the gp120. In general, helices are the most predictable of secondary structures. Turns are weakly predicted following the V2 loop (in the 2nd definable block) and twice within the V3 loop, as we have come to expect. A stronger prediction is for the second turn that follows the third helical region, which separate V3 and V4; this "hinge" coincides with two of the four highly conserved CD4 contact residues and is bordered by the 3rd definable block. The two other contact residues for CD4 interaction occur in the 4th block, which includes the 4th helical region. An examination of predictions for several HIV-1 V3 loop sequences of different subtypes suggests that the Levin, DPM and SOPM methods most consistently predict the putative type II beta turn at the crest of the V3 loop. Further structural analysis of V3 loops is provided by Catasti and Gupta in Part III.


1 10 20 30 40 50 60
most-likely MRVKGIRRNYQHLWRWGILLLGMLMICSAAENLWVTVYYGVPVWKEATTTLFCASDAKAY
Gibrat method EEEEEEEEEHHEHEHHHHHHHHHHHHHHHCCCEEEEEECCCEEECHCCHHHHHHCCHCCC
Levin method ESESSCECCSTTTCTCHHHHHHEHEEECCCTTCEEEEEECCCCCCHCCHEEECCCCCHCH
DPM method CCEECCCCCCCEEEEEEEEEEEEEEECHHHHEEEEEEECECEEHHHCEEEEEHCCHCCC
SOPMA predict HHTTTCCHHHCTEECCCCEEHCCCHHHHTCCCCHEEEEECCCTTCCCCCCCCCCCHHHHH
PhD method CCCCCHHHHHHHHHHHHHHHHHHHHHEECCCCEEEEEEEEEECCCCCCEEEEEECCCCCC
Consensus -CE-C----H---E--HHHHHH-HHHE--CCCEEEEEEECCC--CHCC-EEE--CCHCCC
--- alpha-helix 1 -- turn?
most-likely DTEVHNVWATHACVPTDPNPQEIVLENVTENFNMWKNNMVEQMHEDIISLWDQSLKPCVK
Gibrat method HHHHHEHEHEECECCCCCCCCCHEEEEHHHHHHHHHHHHCHHHHHHHHHHHHHHHCCCCCCC
Levin method HHHHHHHHHHCCECCCCCSCSCEEEHHHHHHHHHHHHHHHHHHHHHHHHHHCHTHCCTTCC
DPM method CCCEEEEEHECCEECCCCCCCCHEEEEEECCCCTCCCCCEEHHHHHCEEEEECCCCCCEEC
SOPMA predict TTTHHHHHHCCEEECCCCCCCEEEEEEEEEHHHCTTCHHHHHHHHCEEECCCCCHHHHHH
PhD method CCCCCCCCCCCCCCCCCCCCEEEEECCHHHHHHHHCHHHHHHHHHHHHHHHCCCCCEE
Consensus ---HH-H-H-CCECCCCCCCC-EEEE--HHHHHHHHCHHHHHHHHHHHH-H-CCCC--C
--> V1 loop <-- -->
most-likely LTPLCVTLNCTDVNATNTNNTTNTTKIDMINETSSCIRQDNCTGLEKGEIKNCSFNITTE
Gibrat method CCCCEECECCCCECCEECCCCCCCEEEEEECCCCCECCCCCCCCCCCCCCEEECCCCCCHH
Levin method CCCEEEEEECECCCHTCTTCCCCCEEEEEECCCCSECCCSTCCCCCTTCCHHCCHHEEHH
DPM method ECCEEEEECCCCECCCCCCCCCCCCCCECEECCCCCEECCCTCCCCTCCCECCTCECEEEE
SOPMA predict HCCTEEEECCEECCCCCTTCCCCCCEEEECCCCCCCCCCCCCCEEEETTEEEEECCHEEH
PhD method CCCCEEEEEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCECCEEEECCCC
Consensus CCC-EEEECCCCCCCCCCCCCCCCEEEEECCCCCECCCCCCCCCCCCCE---C-C-E-H
V2 loop <-- hinge?
most-likely IRDKKQKEYALFYKLDVVPIDNNNTSYRLINCNTSVITQACPKVSFEPIPIHYCAPAGFA
Gibrat method HHHHHHHHHHHHHHHHHEEEECCCCCCEEEEECCCCCCCECCCCCCCCCCCCHCECCCCCHH
Levin method CHCSSHHHHHHHCECCECCCCTCCSEEEEECTCCCEEEEHCCTCCTCSCEEEECCCTSCE
DPM method ECCCCTCHHHEEEEECEECECTTTCCEEEECCCCEEEECCCCCECECCCCCCCCCCCCCEH
SOPMA predict HHHCCHHHHHHHHHHHHEEECCCCCCCEEECCCCCCCEEEEEEEEECCCCEEEECCHHHH
PhD method CCCHHHHHHHHEECCCEEEECCCCCCEEEEECCCCCEEEECCCCCCCCCEEEECCCCEE
Consensus -HCCCHHHHHHH---CEEEECCCCCCEEEECCCCCEEEECCCCCC-CCCCEEECCCCC-H
-- alpha helix 2
most-likely ILKCNDKKFNGTGPCKNVSTVQCTHGIKPVVSTQLLLNGSLAEEEIVIRSENFTDNAKTI
Gibrat method HHHHCCCCCCCCCCCCCCCEEEECCCCCCEEEHHHHHCCCHHHHHEEEHHHCCCHCHHHE
Levin method EEESCCTSCCTCCCCCCCEEECCCCECCCECETCEEEHHCCCTTCEEEECHCCCHCCSCE
DPM method EECTTTCCCTTCTCCCTEEEEEEECCECCEEEEEEECTCCCHHHHEEEECCCCCCCCCEE
SOPMA predict HHHHTCCCCCCCCCCCCCCHHECCTTCCCEEEEEEHHHHHHTTHHHEEHHHHHHHHHHHE
PhD method EEEECCCCCCCCCCCCCEEEEEECCCCCEEEEEEEECCCCCHHHEEEEEEEECCCCCEEE
Consensus EE--CCCCCCCCCCCCCEEEEECCCCCEEEEEE--CCCHHHHEEEE-HCCCHCC--E
V3 loop
--> turn? turn? <-- -- alpha
most-likely IVQLNESVEINCTRPNNNTRKSITIGPGQAFYATGDIIGDIRQAHCNISGAKWNETLQQV
Gibrat method EEEECCCCEEEECCCCCCCCEEEEECCCCEEEECCCHCEEEEEHEECCCCCCHHHHHHHH
Levin method EEEECSHHCEECCCCTCCSCSCEECCTTCCEEETSECEHECCHCCCCHTTCCCHHHHHHH
DPM method EEEECCEEEECCCCCTCTCCCEEEECCCCCECCCCCEECCECHHCCCECTCCCCCCEHEE
SOPMA predict EEEECCCHECCCCCCCCCCCCCEEECCCCTTECCCCCCCCCEEHHEEECCCCCHHHHHHHH
PhD method EEEECCEEEEEECCCCCCCCEEEEEECCCEEEEECCCCCCCCHHHCCHHHHCHHHHHHHHH
Consensus EEEECC--EEECCCCCCCCCCCEEEECCCC-EEECCCCCCCE-HHC-C-CCCCHHHHHHHH
CD4 CD4
helix 3 -- hinge?
most-likely AKKLREQFGNKTIIFNQSSGGDPEITTHSFNCGGEFFYCNTTQLFNSTWNNGTWNSTESN
Gibrat method HHHHHHHHCCCEEEEECCCCCCCEEEEECECCCCEEEEECCEEEEEEEEECCCEEEECCC
Levin method HHHHHHHTTTSEEEECCCSCCCECCECSTCCTCCCEEECCTHHECCTCTTTCCCCTCCT
DPM method HHHHHHCTCCCEEEECTTTTCTCCCCECCCTTCCCEEEECCCCEECCCCTTCCCCCCTTC
SOPMA predict HHHHHHTTTCCEEEEETTTCCCTEEEEEEEEEECCEEEEEEEEECTTCCTTTCCCCCCC
PhD method HHHHHHHCCCCEEEEECCCCCCCEEEEEEEECCCEEEEEEEHHHHCCCCCCCCCCCCCCC
Consensus HHHHHHHTCCCEEEEECCCCCCCEEEEECE-CCCCEEEECC--EECC-C-TCCCCCCCCC
CD4 CD4
-- alpha helix 4 --
most-likely DTITLPCRIKQIINMWQEVGKAMYAPPIEGQITCSSNITGLLLTRDGGDNNSTNETFRPG
Gibrat method CCEEEHHCCHCEHHCHHHHCCCCCCCCCCCEEEECCCCCEEEEEECCCCCCCCCCEEECCC
Levin method SCEEEEECHHHHHHHHHHHTSCECCSCCCTCCCCCTTCCEEEEEETCCCCCTTCCCCCCT
DPM method CCECCCCCEEEEEEEECHECCHCCCCCCCCCCCEECCCCCECCEEECCCTTTTTTCCCCCCTT
SOPMA predict CCEEEECCHHHEEEHHHHHTEEEECCTHCCCCCCCCCCCHEEHCCCCCCCCCCCCCCCCC
PhD method CCCCCCHHHHHHHHHHHHHHHHCCCCCCCCCEEEEECCCCEEEEEECCCCCCCCCEEECCC
Consensus CCEEE--CHHHEHHHHHHH-C-CCCCCCCCCEECCCCCCEEEEEECCCCCCCCCCCCCCC

```
                    1       10        20        30        40        50        60
                    |       |         |         |         |         |         |
                        --alpha helix 5--                 gp120/gp41 fusion domain
most-likely         GGDMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVQREKRAVGLGAVFLGFLGAAGSTMG
Gibrat method       CCCCCCECHHHHHHHHHEHEHCHHHHCCHHHHHHHHHHHHHHHHHHHHHHEEEECHCCCEEE
Levin method        CCCCHHHHHHHHHHHCCCECCCCECCCCCHHHHEHHHCCHHHHCHCEEHCCCCCCCCCCC
DPM method          TCTCCCCCCCCHECEEEEEEECCCCCCCCCCHHHEEEEHHHHHECECEEEEEEECTCCCCCC
SOPMA predict       CCCCCCTHHHHHETTEEEEEECCCCCCCCHHHHHHHHHHHHHHHTTCEEEEEEECCCHHHH
PhD method          CCCCHHHHHHHHCCCCEEEEECCCCCCCCCHHHHHHCCCCCCCCCCHHHHHHHHHHHHHH
Consensus           CCCCCC-HHHHH----EEEECCCCCCCCCCHHHHHHHHHHHHHH---EEEEEEC-CCC---
                                  turn? --       extended helix        --
most-likely         AASITLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARVLAVERYLKDQQ
Gibrat method       EHEEEHHHHHHHHEEHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Levin method        CCCCEEECCCCCCCTHCCCCTTHHHHHHHHHHHHHHCEEEEHCCHHTHHHHHHHHHCTTCC
DPM method          CCCEEEEEEHHHEEEEEEECCCCHEHHHHHHHHEEEEEEEEECHHHHHEEHEHHEHCCCC
SOPMA predict       HHHHEEEEHHHCCCCHHCCCHHHHHHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHHHC
PhD method          HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Consensus           -H--EEE-HHHH---HH--CHHHHHHHHHHHHHHHH---HHHHHHHHHHHHHHHHHHC
                        turn?       turn?       --     extended helix
most-likely         LLGIWGCSGKLICTTTVPWNSSWSNKSLTPIWNNMTWMEWEREIDNYTALIYTLLEESQN
Gibrat method       EEEECEECCCEEEEEEECCCCECCCCCEEEEEHCCCHHHHHHHHHHHHHHHEEEHHHHHH
Levin method        HHHCECCTCCEEETEECCCCCCSTTTTECEHCHHHCHHHHHHCHHHHHHHHHHCCCHCCC
DPM method          EECEETTCCCEEEEEEECCCCCCTTCCCCCCCCCCCCEHHHHHHECCCCEEEEEEHHHTCC
SOPMA predict       EEEEEEETTCEEEEEEECCCEECCCCCCCCCCCCHHHHHHHHHHHHHHCCHHHHHHHCCHHH
PhD method          HHHHCCCCEEEEEEEECCCCCCCCCCCCHHHHHHCCCHHHHHHHHHHHHHHHHHHHHHHH
Consensus           EE-EE--CCCEEEEEEECCCCCCCCCC-C--CHCCCHHHHHHHHHHHHHHHH-HHHHHH
                        extended helix and transmembrane region          -- turn?
most-likely         QQEKNEQELLELDKWASLWNWFDITNWLWYIKIFIMIVGGLIGLRIVFAVLSIVNRVRQG
Gibrat method       HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHEEEECCCCCEEEEEEEEEEEEECC
Levin method        CCCHHHHHHHHHHSCCCCHHHHHHHHHHHHHHHHEEEHHTSCCCHHHHEHHHHHHHCCSS
DPM method          CCTCCCHHHHHHCHCHCEECEECEECEEEEEEEEEEEECEEEEEEEEEEEEEEEECCC
SOPMA predict       HHHHHHHHHHHHCCCCCCCCETCCCCCEEEEEEEEEEEECTTCCEEEEEEEEEEEEEEEC
PhD method          HHHHHHHHHHHHCHHHHHHHHHHHHHHHHHHHHHHHHEEEEECEHHHHHHHHHHHHHHHC
Consensus           HHHHHHHHHHHHCHCHC-HHHHHHHHHHHHHHHHEEEEEEC--C-EEEEEEEEEEEE-CC

most-likely         YSPLSFPPGYIQQTHLPAPRGPDRPEGIEEEGGERDRDRSWRLVNGFLALIWDDLRSLCL
Gibrat method       CCCCCCCCCCEEEECCCCCCCCCCCCCCCCCEHCCCCCCHCHHHHHHHHHHHHHHHHHHH
Levin method        CCCCCCCCSSCEECCCCCCCCCCCCCCHHHHTCCCCHHHHHHHHHHHHHHHHHSTSCCHEH
DPM method          CCCCCCCCCCECCCCCCCCCTCCCCCCCCTCCCTCCCCCCEEEECCEEEEEECCCCCEEE
SOPMA predict       CCEEECCCCCEEEEEEEHHTCCCCHHHHHHTTTCCCHTCCCEECCCCEECCCCCCEEEEE
PhD method          CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHH
Consensus           CCCCCCCCCCEEECCCCCCCCCCCCCCC-HCCCCCCHCCC---HHHHHHHHHC-C--HEH

most-likely         FSYHRLRDLLLIVARIVELLGRSSLKGLRRGWEALKYLWNLLQYWSQELKNSAVSLLNAT
Gibrat method       HHHHHHHHHHHHHHHHHHHHCHCCHHCHEHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Levin method        HETHCHHHHHHHHHHHHHHHCTTCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
DPM method          EEECCEHCEEEEEEEEEEEETTCCTCCTTCCHHHHHEEECEECEECCHTCCCCEEEECHC
SOPMA predict       EETTTTHHHHHHHHHHHHHHTCCCHHHHHHHHHHHHHHHHHHHHHHHHHHTTHEEEECTH
PhD method          HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Consensus           HE-H-HHHHHHHHHHHHHHH--CCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

most-likely         AIAVAEGTDRVIEVLQRIGRAILHIPRRIRQGLERALL
Gibrat method       HHHHHCCCHHHHHHHHHHHHHEEECCCHHHHHHHHHH
Levin method        HHHHHTCCCHHHHHHHHHCHHECHCCHSHCTTCHSHHE
DPM method          HEHEHCTCCEEEEEEEEECEEEEEECCEEECCHHHHCC
SOPMA predict       HHHHHCCTTEEEEEECCCCCEEEECCCEHHHHHHTTTT
PhD method          HHHHHHHHHHHHHHHHHHHHHHHHHCHHHHHHHHHHCC
Consensus           HHHHHCCC-HHHHHHHHHCHHEEECCC-HHHHHHHH--
```

The transmembrane portion of envelope contains more helix than does the surface portion, as determined by circular dichroism, and the structure predictions confirm this for the HMM consensus. An extended helix, encompassed by the 7th definable block, immediately preceeds the immunodominant domain and another, encompassed by the 8th block, follows it. Another predictable helix coincides with the membrane spanning domain, as we would expect. A predicted turn associated with the RQGY peptide that ostensibly signals the terminus of the transmembrane region [24] is not supported by the SOPMA suite.

In the future, we hope to refine the HMM analyses of HIVs and SIVs, first, by addressing the overrepresentation problem of HIV-1s and, second, by integrating the structural information with the primary sequence data.

References

[1] Doolittle, R.F. (1987) *Of URFS and ORFS: A Primer on How to Analyze Derived Amino Acid Sequence* University Science Books, Mill Valley, California.

[2] McClure, M.A., Vasi, T.K., and Fitch, W.M. (1994) Comparative analysis of multiple protein-sequence alignment methods. *Mol. Biol. Evol.* **11**: 571–592.

[3] Baldi, P., Chauvin, Y., Hunkapiller, T., and McClure, M.A. (1994) Hidden Markov models of biological primary sequence information. *Proc. Nat. Acad. Sci. U.S.A.* **91**: 1059–1063.

[4] Krogh, A., Brown, M., Mian, I.S., Sjolander, K., and Haussler, D. (1994) Hidden Markov methods in computational biology: applications to protein modeling. *J. Mol. Biol.* **235**: 1501–1531.

[5] Rabiner, L.R. (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77:257–286.

[6] McClure, M.A. and Raman, R. (1995). Parameterization studies of hidden Markov models representing highly divergent protein sequences. Proceedings of the 28th Annual Hawaii International Conference on Sysytem Sciences, pp. 184–193, published by the IEEE Computer Society Press

[7] Luthy, R. and Eisenberg, D. (1992). Protein. In *Sequence Analysis Primer* (eds. M. Gribskov and J. Devereux), pp. 78–82. W.H. Freeman and Company, New York.

[8] Gribskov, M. and Veretnik, S. (1996). Identification of sequence patterns with profile analysis. In *Computer Methods for Macromolecular Sequence Analysis* (ed. R.F. Doolittle). pp. 146–159, Academic Press, Inc., San Diego.

[9] Eddy,S. (1995) User's guide for HMMER; Hidden Markov Models of protein and DNA sequence (version 1.8), Washington University of St. Louis, 660 S. Euclid, Box 8232, St. Louis MO 63110

[10] Eddy,S., Mitchison,G., and Durbin,R. (1995) Maximum discrimination hidden Markov models of sequence consensus. J Comp Biol 2: 9-23.

[11] Wills,C., Farmer, A., and Myers, G. (1996) Rapid sequon evolution in human immunodeficiency virus type 1 relative to human immunodeficiency virus type 2. AIDS Res Human Retro 12:1383-1384.

[12] Henikoff,S. and Henikoff,J.G. (1994) Protein family classification based on searching a database of blocks. Genomics 19:97–107.

[13] Henikoff,J.G. and Henikoff,S. (1996) BLOCKS database and its applications. In *Computer Methods for Macromolecular Sequence Analysis*, ed. R.F. Doolittle, pp.88–105, Academic Press, San Diego.

[14] Lamers,S.L., Sleasman,J.W., and Goodenow,M.M. (1996) A model for the alignment of ENV V1 and V2 hypervariable domains from human and simian immunodeficiency viruses. AIDS Res Human Retro 12: 1169-1178.

[15] Myers, G.(1996) Retroviral Sequences. In *Retroviruses*, ed. by Coffin,J. Hughes,S. and Varmus,H. Appendix I. Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY.

[16] Geourjon, C. and Deleage, G. (1994) SOPM: a self-optimised prediction method for protein secondary structure prediction. *Prot. Eng.* **7**: 157–164.

[17] Geourjon, C. and Deleage, G. (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *CABIOS* **11**: 681–684.

[18] Gibrat, J.-F., Garnier, J., and Robson, B. (1987) Further developments of protein secondary structure prediction using information theory. *J. Mol. Biol.* **198**: 425–443.

[19] Levin, J.M., Robson, B., and Garnier, J. (1986) An algorithm for secondary structure determination in proteins based on sequence similarity. *FEBS* **205**: 303–308.

[20] Rost, B. and Sander, C. (1994) Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins: Structure, Function and Genetics* **19**: 55–72.

[21] Deleage,G. and Roux,B. (1987) An algorithm for protein secondary structure prediction based on class prediction. Prot Eng 1: 289–294.

[22] Gallaher,W.R., Ball,J.M., Garry,R.F., Griffen,M.C., and Montelaro,R.C. (1989) A general model for the transmembrane proteins of HIV and other retroviruses. *AIDS Res Human Retro* **5**:431–440.

[23] Gallaher,W.R., Ball,J.M., Garry,R.F., Martin-Amedee,A.M., and Montelaro,R.C. (1995) A general model for the surface glycoproteins of HIV and other retroviruses. *AIDS Res Human Retro* **11**:191–202.

Global variation in the HIV-1 V3 region

John C. Blouin, Esther A. Guzman and Brian T. Foley

MS K710, Los Alamos National Laboratory, Los Alamos, NM 87545

Introduction

Due to the immunogenicity and functional importance of the V3 loop, there has been a great deal of interest in the V3 region of the envelope protein, resulting in a large international effort to obtain V3 region sequences. This section, which includes sequences taken from 1651 individuals and complete references, provides an overview of the variation of sequences that span this region.

Sequences

To best summarize the spectrum of international HIV-1 variants, only one representative viral sequence was included per infected individual. A complete set of references accompanies the sequence alignments, and nomenclature was preserved from the original papers so individuals and isolates can be clearly identified. HIV1 was deleted from the sequence names in this section, as all sequences included here are HIV1. Included with the references when available are brief descriptions of critical features of the sequences. This includes the health status of the individual from whom the virus was derived, whether or not the virus was cultured, and the year the blood sample was taken.

All sequences are prefaced by a subtype association (see phylogenetic clustering below) and a two letter country code to identify the country that the individual resided in at the time that the blood sample was taken. If the person was a recent immigrant and this information was available, we included the country of origin in the references. The two letter code was developed for Internet (Copyright 1992, Lawrence H. Landwater and the Internet Society), and incorporated here based on a suggestion made by Dr. Francine McCutchan. The key to the country codes follows this introduction. Note that this key has been updated for 1996, with several country codes for eastern european nations added.

Sometimes only one viral sequence was available from a person: a clone from an isolate, or a direct sequence of PCR amplified peripheral blood DNA. For other individuals, up to 80 viral sequences from PCR amplified DNA or RNA from blood samples were available. Consequently, over 8000 sequences are represented by the 1651 included in this section. When two sequences were available from a person, one of the two was randomly selected. When three or more sequences were generated from a person, all available sequences were aligned (without regard to different time points of sampling) and either one representative sequence was chosen, or a consensus of the most common base found in each position in the alignment was generated. If there was a tie (e.g., 10 A's, 10 T's), the top base or amino acid in the alignment was used. If a set of sequences from two or more individuals was epidemiologically linked, and genetically very similar, only one sequence from the set was included, preferably the most recently infected. In the sequence description and references section, the short hand "PCR-direct, peripheral blood DNA" is used to signify that viral DNA was amplified from PBMCs, without culturing, and a single "direct" sequence was obtained from the amplification reaction products. The short hand "Consensus, PCR-clones, peripheral blood DNA" signifies that viral DNA was amplified from PBMCs and a set of clones was generated and sequenced from the PCR amplification products. The cloned sequences were aligned and a consensus was generated. In a handful of cases, a particular gp160 clone from an isolate was shown to be expressed and functional using a vaccinia virus T7 expression system. In these cases, the clone rather than the consensus of all sequences from a particular individual is included.

Phylogenetic clustering

Sequences have been organized according to the phylogenetic subtype association (A-J) of their envelope V3 regions only. The original sequence subtype (A-H) designations were defined based on the phylogenetic relationships determined by using both gag and env genes (when possible), are

approximately genetically equidistant in envelope, and have multiple members. The phylogenetic subtype designations and associations have generally been adopted by the HIV research community, and are now often presented with the publication of new sequences. We have either determined the subtype designations here, if not specified in the original manuscript, or else confirmed the subtype designations of the original manuscripts, and then used the subtypes to organize this section. Generally, confirmations were done by aligning a set HIV-1 V3 region sequences with longer env gene sequences (Part IIIC) that have clear subtype associations, and then using parsimony or neighbor joining trees to determine associations. Some of the shorter gene fragments from this region were given a subtype designation based on Hamming distances, using the similarity function of the MASE program (Faulkner DV, and Jurka J. TIBS 13:321-322 (1988)); these sequences have ".sh" appended to their name to indicate that they were too short for phylogenetic analysis. Parsimony trees were generated using PAUP (David Swofford, Illinois Natural History Survey), and neighbor-joining trees were generated with Kimura distances and a transition to transversion ratio of 1.3 using PHYLIP (Joseph Felsenstein, University of Washington). All available nucleotide sequence information was used for phylogenetic analysis; longer protein sequences were trimmed to be approximately the same length as the majority of the PCR fragments in this region, for the purposes of presentation. Some sequences were difficult to classify, and are included in the "U", or unclassified, section. In addition, it has recently been noted that recombination between HIV-1 occurs when an individual is infected with more than one strain. A meeting was held in Santa Fe, New Mexico in October, 1995 to discuss the implications of recombination and methods for detecting recombinant sequences. Because inter-subtype as well as intra-subtype recombination is known to occur, the subtype designations reported in this section should be interpretted only as pertaining to the V3 region of the envelope gene. For example HIV-1 MAL from Zaire, is known to be recombinant between subtypes A and D, with the V3 loop of env resembling subtype D. D_ZR-MAL is still listed with other subtype D sequences in this study, but may be moved to the U (uncertain) group in the future.

The set of sequences used to help resolve subtype associations included at least two sequences from each subtype (A-H), plus a simian immunodeficiency virus outgroup sequence. The sequences were selected based on being "typical" of the subtype they represent based on phylogenetic analysis. The set has changed as more sequences have accumulated. Thus not all subtype designations were based on the same reference set.

Limitations of phylogenetic analyses

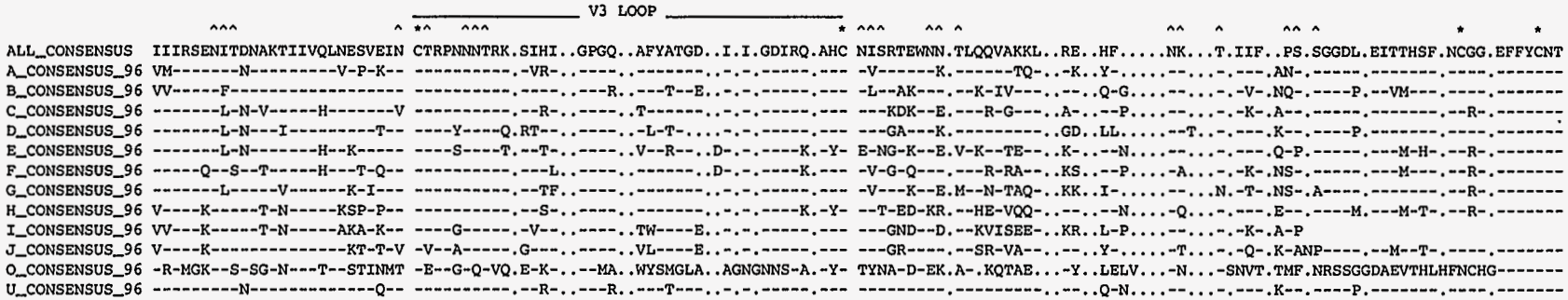
Most of the PCR derived sequences contain a sub-optimal length for phylogenetic analyses, given the level of variability in this region – typically on the order of 250 to 300 nucleotides. Due to this limitation, some of the classifications in this section are uncertain and are our best estimate given the available information. Control studies were performed to compare the phylogenetic clustering of the V3 region using available longer sequences, however, and these studies indicate that our subtype designations based on the V3 region are generally reliable. For 146 sequences, we had an approximately 700 base region of env available representing all of the subtypes A-H. (The limitation in length was due to including the H subtype sequences, which did not cover all of gp120.) After removing positions in the alignment which included gaps, 519 bases were left. When a 298 base V3 region fragment was excised from this set, and neighbor joining trees were constructed using both the 519 base and 298 base long sequences, the phylogenetic subtype designations were consistent in each case. Further, when a subset of longer gp120 sequences was analyzed (92 of the 146), including 935 bases after removing positions in the alignment which included gaps, the subtype designations were again clear in neighbor-joining trees. This indicates that the limited V3 region PCR fragments, which include more than the V3-loop, are generally able to serve as a reliable basis for subtype determination.

Without detailed analysis, genetic recombination between subtypes may obscure phylogenetic relationships between sequences. A characteristic of recombination is an indeterminate place in phylogenetic analyses, and some of the "Uncertain" category sequences may prove to be recombinant genomes upon further inspection. Also, while a subtype designation based on a gene or gene fragment may be correct, recombination events outside the region examined may have occurred. Therefore, care should be taken to not overinterpret the subtype designations. If one is to discuss the subtype

designations of viral isolates based on the data presented here, they should be refer to the designation as "B-like over V3 loop region," rather than as "subtype B".

Limitations of V3 amino acid consensus sequences

The V3 amino acid consensus sequences generated for each subtype have interesting features; however, one should be wary about assuming that any of the consensus sequences may broadly represent their subtype. Certainly many V3 loop variants in each of the subtypes are extremely divergent from the consensus sequences. These divergent forms may have very different biological and immunological characteristics from viruses which are similar to the consensus. Additionally, because of the relatively small sample size of most of the subtypes, consensus sequences can be dominated by a small group of highly similar sequences, which may in turn be a sampling artifact. Hence, these consensus sequences are "evolving" as new sequences from each subtype become available.



Subtype consensus sequences. This V3 region alignment shows a consensus sequence generated for each of the eleven subtypes. The subtype consensus sequences indicate the most common amino acid found in each position among the sequences associated with each subtype. The sequences are aligned to a consensus based on the most common amino acid in the subtype consensus sequences, which approximates a "global" consensus. It was generated in this way (rather than by using all 1651 sequences) to avoid over-representation of the B subtype, which has by far the largest number of available sequences. As is the convention in this compendium, a dash (-) indicates concurrence with the top sequence in the alignment; a period (.) indicates a deletion. The carets show where the N-linked glycosylation sites are found in the consensus. The V3 loop is set off from the surrounding sequence by a space on either side to facilitate viewing. Interesting features of the consensus alignment are: 1) Only in the B subtype is GPGR the most common tip of the V3 loop; globally, GPGQ is more prevalent. 2) A highly conserved N-linked glycosylation site is constitutively absent in the C subtype, proximal to the first cysteine (C) in loop. 3) The D subtype consensus has 34 amino acids from cysteine (C) to cysteine (C) rather than the more common 35; at the point where the deletion occurs, it is not uncommon to find insertions of 2 to 4 amino acids, as can be observed in the sequence alignments. 4) The D subtype has two arginine (R) residues in the V3 loop that are uncharacteristic relative to the other consensus sequences; positively charged amino acids in these positions may result in a syncytia inducing, non-monocytropic phenotype (Fouchier RAM, et al., *J. Virol.* **66**:3183–3187 (1992)). 5) A higher degree of variation is seen in the region just downstream of the V3 loop than within it. This difference is also observed internally among the sequences of the different subtypes. 6) The A, C, G and H consensus sequences have very similar V3 loop sequences.

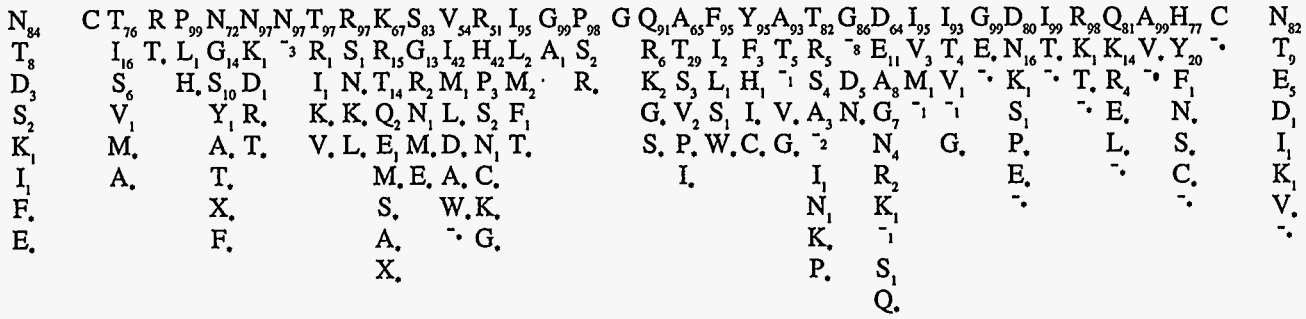
V3 Loop Amino Acids

The following pages present amino acid alignments of the V3 loop, arranged by phylogenetic subtype. For each subtype, the number of sequences used to construct the alignment is indicated. The top line in each alignment represents the consensus sequence for that subtype, where consensus simply means the most common amino acid found in each position among the sequences of the given subtype. The subscripts record the frequency with which that amino acid is observed at that location among members of the subtype. An amino acid which is conserved 100% is shown with no subscript. Directly beneath the most common amino acid in each position are the other amino acids observed in that position, listed from most common to least common. An asterisk (*) subscript means less than 0.5% of the sequences had the indicated amino acid at that location. A dash (-) indicates a gap inserted to maintain the alignment. Percentages were rounded to the nearest whole number.

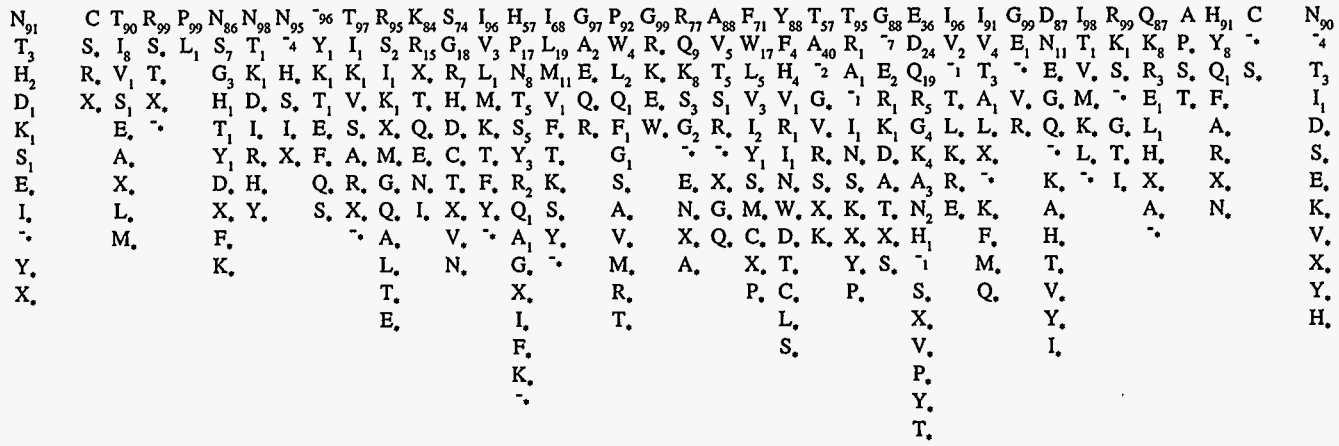
For this year's alignment, the HMMER (version 1.8) hidden Markov model software (Sean Eddy, Dept. of Genetics, Washington U. School of Medicine, St. Louis, MO 63110; eddy@genetics.wustl.edu) was used to objectively align all 1651 sequences. The frequency counts are derived from this alignment. Because each subtypes required different numbers and positions of gaps in order to create the full multiple sequence alignment, some sequences with unusual insertions were trimmed from the HMMER alignment, and a few positions were adjusted by hand, using MASE, prior to printing the full alignment which appears following the country codes description. The sequences which were culled from the alignment after counting frequencies, are appended.

Both the untouched HMMER alignment, and the edited version, will be available via ftp from the LANL HIV database (http://hiv-web.lanl.gov). Questions about these alignments should be directed to (btf@t10.lanl.gov) (505-665-1970).

A subtype (207 sequences)



B subtype (975 sequences)



C subtype (119 sequences)

D subtype (107 sequences)

E subtype (124 sequences)

F subtype (59 sequences)

N_{90}	T_{5}	D_{2}	X_{2}	$-_{2}$				
C_{98}	$-_{2}$							
T_{97}	S_{2}	$-_{2}$						
R_{98}	$-_{2}$							
P_{95}	F_{2}	X_{2}	$-_{2}$					
N_{95}	D_{2}	S_{2}	$-_{2}$					
N_{97}	K_{2}	$-_{2}$						
N_{95}	$-_{5}$							
$-_{97}$	Y_{2}	T_{2}						
T_{95}	I_{3}	$-_{2}$						
R_{98}	$-_{2}$							
K_{92}	R_{5}	X_{2}	$-_{2}$					
S_{86}	G_{8}	R_{2}	X_{2}	$-_{2}$				
I_{98}	$-_{2}$							
H_{71}	R_{10}	Q_{7}	P_{5}	X_{3}	Y_{2}	$-_{2}$		
L_{58}	I_{32}	F_{3}	X_{3}	V_{2}	$-_{2}$			
G_{98}	$-_{2}$							
P_{97}	L_{2}	V_{2}						
G								
Q_{66}	R_{34}							
A_{86}	V_{8}	T_{2}	P_{2}	X_{2}				
F_{93}	I_{3}	L_{2}	X_{2}					
Y_{95}	F_{3}	H_{2}						
A_{66}	T_{27}	X_{3}	R_{2}	S_{2}				
T_{97}	A_{3}							
G_{90}	S_{3}	A_{2}	D_{2}	V_{2}	$-_{2}$			
D_{66}	E_{15}	A_{5}	G_{3}	N_{3}	X_{3}	Q_{2}	K_{2}	
I_{97}	V_{3}							
I_{90}	T_{5}	N_{2}	V_{2}	X_{2}				
G								
D_{93}	N_{7}							
I								
R								
K_{90}	Q_{8}	R_{2}						
A								
H_{83}	Y_{15}	C_{2}						
C								
N_{83}	K_{3}	V_{3}	X_{3}	S_{2}	T_{2}	D_{2}	I_{2}	

G subtype (23 sequences)

N_{52}	T_{17}	I_{13}	V_{13}	M_{4}		
C						
T_{87}	I_{9}	V_{4}				
R						
P						
N_{91}	S_{9}					
N						
N						
T						
R_{96}	I_{4}					
K_{96}	R_{4}					
S_{96}	R_{4}					
I_{96}	K_{4}					
T_{35}	R_{22}	H_{13}	K_{9}	S_{9}	N_{9}	P_{4}
F_{57}	I_{26}	L_{17}				
G_{91}	A_{9}					
P_{91}	T_{9}					
G						
Q_{91}	R_{9}					
A_{78}	T_{13}	V_{9}				
F_{70}	L_{22}	I_{9}				
Y						
A_{96}	T_{4}					
T						
G_{78}	D_{9}	S_{9}	T_{4}			
D_{35}	E_{22}	A_{17}	N_{13}	Q_{9}	G_{4}	
I_{96}	V_{4}					
I_{96}	T_{4}					
G						
D_{70}	N_{30}					
I						
R_{96}	K_{4}					
Q_{96}	K_{4}					
A						
H_{74}	Y_{26}					
C						
N_{96}	K_{4}					

H subtype (2 sequences)

N	
C	
T	
R	
P	
N	
N	
N	
T	
R	
K_{50}	R_{50}
S	
I_{50}	M_{50}
R_{50}	S_{50}
I	
G_{50}	$-_{50}$
I_{50}	$-_{50}$
G	
P_{50}	R_{50}
G	
Q	
A_{50}	T_{50}
F	
H_{50}	Y_{50}
A	
I_{50}	T_{50}
G	
A_{50}	D_{50}
I	
I	
G	
D	
I	
R	
K_{50}	Q_{50}
A	
H_{50}	Y_{50}
C	
N	

I subtype (1 sequence)

N C T R P G N N T R K S V H I G P G Q T W Y A T G E I I G D I R Q A H C N

J subtype (5 sequences)

V_{40}	K_{20}	T_{20}	E_{20}
C			
V_{60}	T_{20}	N_{20}	
R			
P			
A_{60}	N_{20}	Y_{20}	
N			
N			
T			
R			
K_{80}	E_{20}		
G_{80}	S_{20}		
I			
H			
I_{60}	M_{40}		
G			
P			
G			
Q			
V			
L			
Y			
A			
T			
G			
E_{60}	G_{20}	Q_{20}	
I_{60}	V_{40}		
I			
G			
D_{60}	N_{40}		
I			
R			
Q_{60}	E_{40}		
A			
H			
C			
N			

O subtype (13 sequences)

Position	Residues
1	T_{92} A_{8}
2	C
3	E_{31} H_{15} I_{8} M_{8} N_{8} Q_{8} S_{8} T_{8} V_{8}
4	R
5	P_{77} E_{8} K_{8} D_{8}
6	$-_{92}$ E_{8}
7	G_{85} E_{8} Q_{8}
8	N_{46} I_{15} R_{15} M_{8} H_{8} D_{8}
9	$-_{77}$ H_{8} N_{8} D_{8}
10	L_{23} Q_{23} $-_{23}$ G_{15} V_{8} A_{8}
11	$-_{92}$ V_{8}
12	T_{31} $-_{31}$ K_{23} E_{8} S_{8}
13	V_{46} I_{38} $-_{15}$
14	Q_{85} K_{8} $-_{8}$
15	E_{62} D_{8} K_{8} Q_{8} R_{8} S_{8}
16	I_{69} M_{23} V_{8}
17	K_{31} G_{15} R_{15} A_{8} M_{8} Q_{8} H_{8} Y_{8}
18	$-_{92}$ E_{8}
19	$-_{92}$ M_{8}
20	$-_{92}$ K_{8}
21	I_{69} T_{15} S_{8} A_{8}
22	G
23	P
24	M_{77} L_{23}
25	$-_{62}$ S_{23} A_{8} R_{8}
26	A_{54} $-_{31}$ V_{8} R_{8}
27	W_{92} Y_{8}
28	Y_{77} S_{8} V_{8} R_{8}
29	$-_{92}$ G_{8}
30	S
31	M_{85} L_{15}
32	G_{46} A_{23} Q_{8} S_{8} T_{8} $-_{8}$
33	L_{62} I_{23} D_{8} $-_{8}$
34	A_{38} E_{15} G_{15} K_{15} L_{8} S_{8}
35	A_{31} R_{23} G_{15} E_{15} N_{8} K_{8}
36	G_{23} N_{23} E_{15} T_{15} D_{8} S_{8} K_{8}
37	N_{38} S_{15} T_{15} A_{8} G_{8} K_{8} $-_{8}$
38	G_{23} N_{23} T_{15} K_{8} E_{8} S_{8} A_{8} $-_{8}$
39	N_{31} S_{23} K_{8} G_{8} I_{8} T_{8} V_{8} $-_{8}$
40	$-_{31}$ S_{15} E_{15} N_{8} Q_{8} R_{8} A_{8} K_{8}
41	$-_{77}$ N_{15} P_{8}
42	$-_{85}$ S_{8} R_{8}
43	S_{62} T_{23} K_{8} $-_{8}$
44	$-_{92}$ G_{8}
45	$-_{85}$ G_{8} D_{8}
46	$-_{85}$ T_{8} S_{8}
47	R
48	A_{31} I_{15} R_{15} V_{15} S_{8} T_{8} K_{8}
49	A
50	Y_{85} H_{8} F_{8}
51	C
52	N_{23} K_{15} T_{15} M_{8} E_{8} R_{8} D_{8} X_{8} $-_{8}$

U subtype (15 sequences)

Position	Residues
1	N_{80} K_{7} S_{7} T_{7}
2	C
3	T_{93} I_{7}
4	R
5	P
6	N_{47} S_{20} G_{13} Y_{13} A_{7}
7	N_{80} K_{7} S_{7} T_{7}
8	N_{73} $-_{27}$
9	$-_{73}$ I_{7} K_{7} T_{7} D_{7}
10	$-_{87}$ R_{7} K_{7}
11	$-_{87}$ N_{7} K_{7}
12	T_{60} I_{33} K_{7}
13	$-_{93}$ K_{7}
14	R
15	K_{33} R_{27} Q_{20} E_{7} T_{7} X_{7}
16	S_{60} G_{13} R_{7} I_{7} T_{7} X_{7}
17	I_{53} T_{27} M_{13} V_{7}
18	R_{47} H_{40} P_{7} N_{7}
19	I_{80} F_{13} L_{7}
20	G
21	P_{80} L_{13} S_{7}
22	G
23	Q_{47} R_{47} K_{7}
24	A_{73} T_{13} V_{13}
25	F_{73} L_{13} I_{7} Y_{7}
26	Y_{80} F_{20}
27	T_{47} A_{40} K_{7} $-_{7}$
28	T_{73} N_{13} S_{7} K_{7}
29	$-_{47}$ G_{40} X_{7} R_{7}
30	$-_{93}$ G_{7}
31	$-_{93}$ V_{7}
32	$-_{93}$ E_{7}
33	$-_{93}$ G_{7}
34	$-_{93}$ K_{7}
35	$-_{93}$ K_{7}
36	$-_{93}$ G_{7}
37	$-_{93}$ P_{7}
38	D_{47} G_{13} $-_{13}$ K_{7} N_{7} X_{7} E_{7}
39	$-_{93}$ P_{7}
40	I_{67} M_{13} $-_{13}$ V_{7}
41	I_{60} T_{13} K_{7} G_{7} V_{7} $-_{7}$
42	G_{93} $-_{7}$
43	D_{80} $-_{13}$ N_{7}
44	I_{87} $-_{13}$
45	R_{87} $-_{13}$
46	Q_{73} K_{20} $-_{7}$
47	A_{93} $-_{7}$
48	H_{67} Y_{27} Q_{7}
49	C
50	N_{73} K_{7} I_{7} S_{7} T_{7}

Summary of variations in the tetrameric tip of the V3 loop. This table is a tally of the different tetramers observed in the 1651 individuals analyzed. This tip is thought to form a turn, and is the focal point of the potent neutralizing antibody epitopes that have been mapped to the V3 loop, as well as of T cell epitopes. Each column shows the number of occurrences of a given tetramer in either the entire 976 sequences (combined), or in subsets consisting of subtypes A–O, and the unclassified sequences (U). Underneath the column heading is the number of sequences in each category. The most common form found in each subtype is highlighted in bold lettering. In the B subtype, GPGR is the predominant form, however globally GPGQ is more common.

	Combined	A	B	C	D	E	F	G	H	I	J	O	U
Totals	1651	208	975	119	107	124	59	23	2	1	5	13	15
GPGR	711	10	**643**		17	18	17						**6**
GPGQ	590	**184**	91	**119**	**31**	**95**	**39**	**19**	**1**	**1**	**5**		5
GPGK	81	4	76										1
GWGR	34		34										
GPGS	26	1	25										
GPGG	22	2	20										
GLGQ	21				20								1
GLGR	19		15		1	1	1						1
APGR	17	1	16										
GSGQ	16	3			12								1
GQGQ	13		1		12								
GQGR	9		6		1	2							
GPMA	8											**8**	
GPGH	6					6							
GFGR	6		6										
GTGQ	5				5								
GRGQ	5	1			3				1				
GVGR	4		2		1		1						
GSGR	4	1	3										
GPRR	3		3										
GPGA	3		3										
EPGR	3		3										
APGS	3		3										
APGQ	3	1						2					
GTGR	2							2					
GPKR	2		2										
GMGR	2		2										
GGGR	2		2										
GAGR	2		2										
AGGR	2		2										
GLGS	1		1										
GTGG	1		1										
GPMR	1											1	
GLRQ	1				1								
AQGR	1				1								
GPMS	1											1	
GPLR	1											1	
GARR	1		1										
GGGQ	1					1							
GPWG	1		1										
GIGQ	1				1								
RPGR	1		1										

	Combined	A	B	C	D	E	F	G	H	I	J	O	U
Totals	1651	208	975	119	107	124	59	23	2	1	5	13	15
GRGR	1		1										
GPGR	1		1										
GQGI	1					1							
GPLS	1											1	
GPWG	1		1										
GPGN	1		1										
GPGX	1		1										
GPGE	1		1										
GPEK	1		1										
GLGK	1		1										
GARR	1		1										
AWGR	1		1										
APGG	1		1										
AGGK	1		1										
AQGR	1				1								
GIGQ	1				1								
GGRA	1				1								
*PGR	1						1						

Summary of variations in the octameric tip of the V3 loop. This table is a tally of the different octamers observed in the 1651 individuals analyzed. This table is structured the same as the tetramer table on the previous pages. Amino acid changes proximal to the tip can influence the specificity of anti-V3 neutralizing antibodies. The forms that were found only once in the data set are not shown here, to save space, and are summarized in a row labeled "unique."

	Combined	A	B	C	D	E	F	G	H	I	J	O	U
Totals	1651	208	975	119	107	124	59	23	2	1	5	13	15
HIGPGRAF	279	3	**269**		2		3						**2**
RIGPGQTF	136	46		**82**	1	2	1	**3**					1
RIGPGQAF	75	45		27		1		1					1
PIGPGRAF	71	1	69				1						
HIGPGQAF	62	**56**	2				4						
NIGPGRAF	59		58										1
HLGPGQAW	44		44										
TIGPGQVF	39					**39**							
PLGPGQAW	31		31										
HLGPGQAF	31	2	1				**28**						
SIGPGRAF	26	1	25										
HIGPGKAF	25	2	23										
RIGPGQVF	23	3		2		16	1						1
TIGPGRAF	19		19										
HIGPGQAL	16		2		**13**			1					
YIGPGRAF	15		15										
PIGLGQAL	14				13								1
HMGWGRAF	14		14										
HIAPGRAF	14	1	13										
PLGPGRAW	11		11										
HLGPGRAW	10		10										
HIGPGSAF	10		10										
HIGPGRAY	10				10								
TMGPGQVF	9					9							
PMGPGRAF	9		9										
HMGPGRAF	9	1	8										
RIGPGRVF	8					8							
PIGPGQAF	7	3			1	2		1					
HMGWGRTF	7		7										
HIGPGRTF	7		7										
TIGPGRVF	6		1			5							
QIGPGRAF	6		5				1						
PMGPGKAF	6		6										
PLGPGKAW	6		6										
NIGPGRAW	6		6										
HLGWGRAF	6		6										
HIGPGRAV	6		6										
YLGPGRAF	5		5										
RFGPGQAF	5	1					1	1					2
PIGPGKAF	5		5										
NMGPGRAF	5		5										

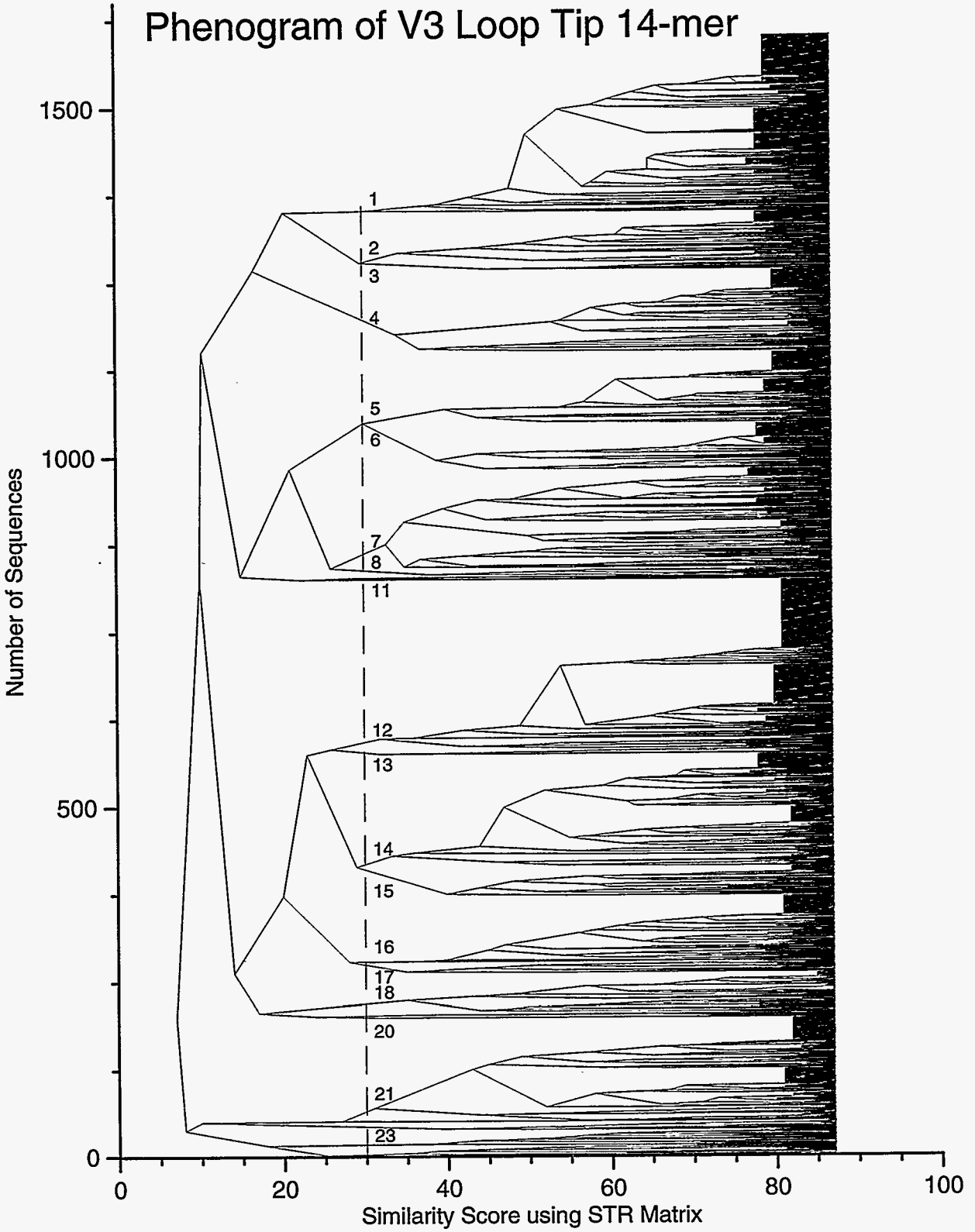
	Combined	A	B	C	D	E	F	G	H	I	J	O	U
Totals	1651	208	975	119	107	124	59	23	2	1	5	13	15
NIGPGQVF	5					5							
HLGPGRAF	5	1	1				2						1
HLGPGGAF	5		5										
HIGSGQAL	5				5								
HIGPGRAW	5		5										
HIGPGRAL	5		2		1		1						1
HIGPGRAI	5		5										
HIGPGGAF	5		5										
SIGQGQAL	4				4								
SIGPGQAF	4	3							1				
PIGPGRAW	4		4										
PIGPGQVF	4					4							
KIGPMAWY	4											4	
HMGPGKAF	4		4										
HMGLGRAF	4		4										
HLGPGQAL	4		2		2								
HIGPGRVF	4		3				1						
HIGPGRSF	4	2	2										
HIGPGQVF	4	1				3							
HIGPGQTF	4	4											
HIGPGQAI	4	2			2								
YIGPGRAV	3		3										
YIGPGRAS	3		3										
TMGPGRVY	3		3										
TMGPGRVW	3		3										
TLGPGRVY	3		3										
TIGPGRVY	3		1			2							
TFGPGQAF	3							3					
SLGPGRAW	3		3										
SIGPGRAW	3		3										
RIGPGQTL	3			2	1								
RIGPGQSF	3	3											
PLGPGRAF	3		1				2						
NIGPGQAF	3	2			1								
HMGPGKTF	3	1	2										
HLGQGRAW	3		3										
HIGTGQAL	3				3								
HIGSGQAY	3				3								
HIGPGQVL	3										3		
HIGPGQAW	3		3										
GIGPGQTF	3	2		1									
AIGPGQVF	3					3							
XIGPGRAF	2		2										
TRGPGHVF	2					2							
TMGPGRVL	2		2										
TMGPGKVF	2		2										
TMGPGHVF	2					2							
TLGPGQAF	2							2					
TIGPGQVL	2					2							

	Combined	A	B	C	D	E	F	G	H	I	J	O	U
Totals	1651	208	975	119	107	124	59	23	2	1	5	13	15
TIGPGQIF	2					2							
SMGPGRAF	2		2										
SLGPGKAW	2		2										
SIGQGRVL	2					2							
SIGQGQTL	2				2								
SIGPGRVW	2		2										
SIGLGQAL	2				2								
SFGPGQAF	2							2					
RIGPMAWY	2											2	
RIGPGSAF	2		2										
RIGPGRVI	2						2						
RIGPGRTF	2		1										1
RIGPGRAV	2		2										
RIGPGRAF	2		2										
RIGPGQAL	2				2								
QLGPGRAW	2		2										
PLGPGRVW	2		2										
PIGSGQAL	2				2								
PIGRGQAL	2				2								
PIGLGQAY	2				2								
PIAPGSAW	2		2										
KIGPGQTF	2	1			1								
KFGTGRVL	2							2					
HVGPGQAF	2				1		1						
HMGPGRAL	2				2								
HMGPGQVL	2										2		
HMGPGGAF	2		2										
HLGPGKAW	2		2										
HLGPGKAF	2		2										
HLGLGRAF	2		2										
HLGFGRAL	2		2										
HIGSGRAF	2	1	1										
HIGSGQAI	2				1								1
HIGPGSAL	2		2										
HIGPGQAY	2				2								
HIGPGKVF	2		2										
HIGLGRAY	2				1								1
HIGGGRTL	2		2										
HIEPGRAF	2		2										
HFGPGQAL	2				1			1					
GIGPGRTV	2		2										
AIGPGRTV	2		2										
UNIQUE	189	26	93	3	26	8	10	11	2	3	7		

Phenogram of the 14 amino acid tip of the V3 loop.

The immunogenic tip of the V3 loop has been the focus of a large number of serological studies. Here we have organized the 14 amino acid tip of the V3 loop of each sequence according to their amino acid similarities. The four amino acid tip of the loop (the GPGR motif) plus the 5 amino acids on either side were clustered using maximum linkage based on pairwise distance scores using the STR matrix (S. Henikoff and J. G. Henikoff, *Proteins: Structure, Function and Genetics* **17**: 49–61 (1993)). The resulting phenogram of all 1651 is shown, with related sequences grouped into 25 clusters. The y-axis represents the number of sequences included in the phenogram, and the x-axis indicates the level of similarity between amino acid sequences; the higher the score at the branch junction, the more similar the sequences. The 25 clusters were selected such that the sequences within a cluster all are related by scores greater than 30. The alignment of these sequences corresponding to the phenogram is included on pages following the figure. The clusters are numbered on the phenogram and in the alignment. The order of the sequences in the alignment corresponds to the order that the sequences appear in the phenogram, from top to bottom.

There are two things worth noting in the alignment. First is the intermixing of amino acid sequences from different clades—often sequences from phylogenetically distinct subtypes have identical or nearly identical sequences in this region. This is easily noted in clusters 1, 2 and 6 which contain A and C subtypes. Second is the diversity of the D subtype sequences. The two letter country codes can be used to follow geographic associations of similar sequences, or alternatively, the range of variation of a given subtype in a country of interest. When this type of phenogram is computed using the entire 35 amino acids of the V3 loop, a very similar result is obtained (B.T.M. Korber et al., *J. Virol.* **68**:6730–6744 (1994)). When a phenogram is computed using 35 amino acids flanking the V3-loop (20 from the amino side and 15 from the carboxy side), or 14 amino acids (7 from each side of the V3 loop) the resulting phenogram is quite different, with A and C subtype sequences residing on seperate branches.


Phenogram of V3 Loop Tip 14-mer
Number of Sequences
1500
1000
500
0
1
2
3
4
5
6
7
8
11
12
13
14
15
16
17
18
20
21
23
0
20
40
60
80
100
Similarity Score using STR Matrix

ALL_CONSENSUS	KSIHIGPGRAFYAT
	1
G_NL1.LR94018	---R----QT----
G_TW1.267	---R----QT----
G_GA1.G98	---R----QT----
D_TZ3.1585	---R----QT----
C_ZW.2647	---R----QT----
C_ZA2.Dlu	---R----QT----
C_RU.RUS2A	---R----QT----
C_RU.RUS20A	---R----QT----
C_RU.RUS1A	---R----QT----
C_RU.RUS13A	---R----QT----
C_NO1.V3N13	---R----QT----
C_MY.9214082	---R----QT----
C_MW2.D3MA960	---R----QT----
C_MW1.6518	---R----QT----
C_MW1.12213	---R----QT----
C_IN4.CMCH8	---R----QT----
C_IN4.CMCH7	---R----QT----
C_IN4.CMCH6	---R----QT----
C_IN4.CMCH5	---R----QT----
C_IN4.CMCH4	---R----QT----
C_IN4.CMCH37	---R----QT----
C_IN4.CMCH36	---R----QT----
C_IN4.CMCH33	---R----QT----
C_IN4.CMCH32	---R----QT----
C_IN4.CMCH3	---R----QT----
C_IN4.CMCH29	---R----QT----
C_IN4.CMCH27	---R----QT----
C_IN4.CMCH26	---R----QT----
C_IN4.CMCH23	---R----QT----
C_IN4.CMCH22	---R----QT----
C_IN4.CMCH20	---R----QT----
C_IN4.CMCH2	---R----QT----
C_IN4.CMCH19	---R----QT----
C_IN4.CMCH15	---R----QT----
C_IN4.CMCH13	---R----QT----
C_IN4.CMCH12	---R----QT----
C_IN4.CMCH11	---R----QT----
C_IN4.CMCH10	---R----QT----
C_IN4.CMCH1	---R----QT----
C_IN3.IND8	---R----QT----
C_IN3.IND7	---R----QT----
C_IN3.IND6	---R----QT----
C_IN3.IND4	---R----QT----
C_IN3.IND3	---R----QT----
C_IN3.IND2	---R----QT----
C_IN3.IND	---R----QT----
C_IN2.D868	---R----QT----
C_IN2.D744	---R----QT----
C_IN2.D1024	---R----QT----
C_IN1.D760	---R----QT----
C_IN1.D757	---R----QT----
C_IN1.D1044	---R----QT----
C_ET2.E1439	---R----QT----
C_CY.HO021	---R----QT----
C_CONSENSUS_96	---R----QT----
C_BY1.BLR8A	---R----QT----
C_BY1.BLR5A	---R----QT----
C_BR2.91BR015	---R----QT----
A_RW2.1831cons	---R----QT----
A_NL1.ZR93033	---R----QT----
A_CI2.CI20	---R----QT----
C_RU.YAN4	E--R----QT----
C_UG.UG268	E--R----QT----
C_GA1.G134	E--R----QT----
C_ET2.E6613	E--R----QT----
C_ET2.E6209	E--R----QT----
C_ET2.2220	E--R----QT----
C_DJ1.DJ259	E--R----QT----
C_BU1.91BU006	E--R----QT----
C_DJ1.DJ373	Q--R----QT----
C_MW.SM750	---R----QK----
A_IN1.CMCH9	---K----QT----
A_UG2.78	R-VR----QT----
A_ZR3.L414	R-VR----QT----
A_UG1.W2UG037	R-VR----QT----
A_UG.92UG037	R-VR----QT----
A_NL1.GH94014	R-VR----QT----
A_CM1.CA15	R-VR----QT----
A_BJ1.256	R-VR----QT----
C_BU1.91BU003	R--R----QT----
C_IN2.D808	--TR----QT----
C_NL1.UN93050	--MR----QT----
C_SN.SE364	--MR----QT----
C_ET2.E72150	--MR----QT----
C_ET2.E2564	--MR----QT----
C_ET2.E1325	--MR----QT----
C_ET2.E1320	--MR----QT----
A_CF1.4023	R-MR----QT----
A_CI2.CI329	--LR----QT----
C_IN2.D766	--LR----QT----
C_MW1.12233	--MR----QP----
C_ZM1.ZAM20	---R----QT-F--
F_CY.HO441	---R----QT-F--
C_RW1.134cons	---R----QT-F--
C_ZA.ZA514	--VR----QT-F--
C_ZA.ZA517	---R----Q--F--
U_KE.K124	---R----Q--F--
C_GM.GM3	---R----Q--F--
A_RW2.561Wcons	--VR----Q--F--
B_BR3.HRJ626	---R----S-----
B_BR3.RJ626	---R----S-----
B_SK1.BTS22	---R----ST----
C_ET2.E7827	---R----QTL---
C_RW1.566cons	---R----QTL---
C_BU1.91BU001	---R----QIL---

E_CF.4081	--VR----QT----
U_CF1.408	--VR----QT----
C_SO.1574	--VR----QT----
C_NO1.V3N19	--VR----QT----
C_MW1.6508	--VR----QT----
C_MW1.12229	--VR----QT----
C_MW1.12225	--VR----QT----
A_ZR1.6657	--VR----QT----
A_ZR1.6563	--VR----QT----
A_UG2.117	--VR----QT----
A_RW4.564cons	--VR----QT----
A_RW.564C	--VR----QT----
A_RU1.IVA6	--VR----QT----
A_NL1.RW8903	--VR----QT----
A_NL1.GH9152	--VR----QT----
A_GA1.VI1076	--VR----QT----
A_GA1.LBV2310	--VR----QT----
A_DJ.DJ264	--VR----QT----
A_DJ.DJ263	--VR----QT----
A_DJ.DJ258	--VR----QT----
A_CM1.CA2	--VR----QT----
A_CM1.CA19	--VR----QT----
A_CI2.CI45	--VR----QT----
A_CI1.4con	--VR----QT----
A_BJ1.44	--VR----QT----
A_BJ1.36	--VR----QT----
A_BJ1.33	--VR----QT----
A_BJ1.253	--VR----QT----
A_BJ1.252	--VR----QT----
A_BJ1.234	--VR----QT----
A_BJ1.218	--VR----QT----
A_BJ1.1	--VR----QT----
A_RW4.439cons	--VR----QT---S
A_CM1.CA17	--VR----QT---A
A_CI1.46con	--VR----GQTFYA
C_SO.SM145	--VR----QT--TN
C_ZM1.ZAM18	---R----Q-----
G_NG1.G3	---R----Q-----
C_ZA2.GOM	---R----Q-----
C_TW1.252	---R----Q-----
C_MY.9214083	---R----Q-----
C_MW1.1227	---R----Q-----
C_MW1.12215	---R----Q-----
C_MW1.12203	---R----Q-----
C_MW1.12199	---R----Q-----
C_KE.NA113	---R----Q-----
C_BY1.BLR9A	---R----Q-----
C_BR1.HSP203	---R----Q-----
C_BR.W2BR025	---R----Q-----
A_ZR1.6653	---R----Q-----
A_UG5.94UG019	---R----Q-----
A_UG5.94UG009	---R----Q-----
A_NL1.KE9135	---R----Q-----
A_CF1.4055	---R----Q-----
A_CF1.4054	---R----Q-----
A_CF1.4033	---R----Q-----
A_CA.HWCL1	---R----Q-----
A_BJ1.41	---R----Q-----
A_BJ1.193	---R----Q-----
ALL_CONSENSUS	---R----Q-----
A_RW2.1831Wcons	---R----Q----S
A_RW1.W2RW026	R--R----Q-----
C_NO1.V3N14	R--R----Q-----
A_NL1.GH9023	R--R----Q-----
A_KE.KEN967	R--R----Q-----
A_RW3.PVP5	R--R----Q----S
A_ZR1.6655.sh	---R----QV----
F_CM.CA4	---R----QV----
C_MW1.12205	Q--R----QV----
A_KE.K89	M--R----Q-----
A_KE1.NA112	S--R----Q-----
A_UG5.94UG012	--VR----Q-----
C_UK1.00513	--VR----Q-----
A_UG1.W2UG03	--VR----Q-----
A_TZ2.017	--VR----Q-----
A_TZ2.016	--VR----Q-----
A_RW2.1613Wcons	--VR----Q-----
A_RW1.W2RW02	--VR----Q-----
A_RW.SF1703	--VR----Q-----
A_CONSENSUS_96	--VR----Q-----
A_CI2.CI330	--VR----Q-----
A_CI2.CI3	--VR----Q-----
A_ZR1.6557.sh	Q-VR----Q-----
C_MW1.6512	Q-VR----Q-----
A_UG.964	E-VR----Q-----
C_UG.45	E-VR----Q-----
C_BU1.91BU004	E--R----Q-----
C_MW1.12209	Q--R----Q--F--
C_NO1.V3N17	Q--R----Q-----
A_RW4.618cons	T--R----Q--F--
C_BU1.91BU005	T--R----Q-----
C_BU1.91BU008	T--R----Q-----
A_NL1.TZ925825	T--R----Q-----
A_KE.KEN98	T--R----Q-----
A_KE.KEN977	T--R----Q-----
A_BY.BLR10A	T--R----Q-----
A_UG4.UG273	T-VR----Q-----
A_ZR1.6663	T-VR----Q-----
A_UG2.82	T-VR----Q-----
A_RW2.1701cons	T-MR----Q-----
A_RW1.W2RW024	T-VR----QT----
A_RW1.W2RW025	T-VR----QT----
A_KE1.NA117	T-VR----QT----
A_KE.KEN88	T--R----QT----
A_UG4.UG275	T-VR----QS----
A_KE.KEN985	---R----QT---G
A_TZ3.1574	---R----Q----G
A_BJ1.23	---R----QS---A

```
A_RW3.PVP2          E--R----QT---A
C_BU1.91BU002       E--R----QT---H
C_GA1.LBV105        E--R----Q----P
A_RW1.W2RW008       T--R----QS-H--
C_ZA2.BooyD         ---R----Q--H--
E_CF1.4039          T-VR----QV--K-
E_CF1.4084          T-VR----QV--K-
E_CF1.403           T-VR----QV--K-
E_CF1.4017          T-VR----QV--K-
E_TH9.100           T--R----QV--K-
E_CF1.1697          T-AR----QV--K-
E_CF2.ELO           T-AR----QV--K-
E_CF2.MBA           T-VR----QV--G-
E_TW1.396           A-LR----QV--T-
C_MW2.D3MA959       R--R----QV---N
U_NL.RW94028        ---R----QV---N
A_ZR1.6571.sh       ---R----QV--TN
U_ZR.Z3             Q--R----KV---K
A_CF1.4050          R-VR----Q-----
A_CM1.CA18          R-VR----Q-----
C_NL1.NL94024       R-VR----Q---T-
A_CM1.CA22          R-VR----Q-I---
A_CI1.17con         -GVR----QT----
C_IN1.D747          -GVR----QT----
A_NL1.GH9299        -GVR----QV----
A_RW3.PVP           -GVR----Q-W--R
A_KE.KEN965         T--R----Q---GR
E_CM.CA10           T-SR----Q---KI
A_KE.KEN979         T--C----QT---G
A_KE1.NA115         T--S----K--F-S
A_UG2.11            T-V-----Q----R
A_UG2.115           --------Q----R
A_KE.KEN29          TG------Q----R
A_CI2.CI327         -GV-----Q----R
A_GH.D687           -NV-----Q----R
A_ZR1.6559          QGV-----QV---R
D_ZA.ZA500          QNTK----QT-F-R
A_CM1.CA7           R--R--S-QTS---

                2

C_ZA.NOF            -R-RV---QTV---
A_UG2.124           --V-----Q-----
A_ZR1.6649          --V-----Q-----
A_UG2.116           --V-----Q-----
A_TZ3.1576          --V-----Q-----
A_TZ.TAN8           --V-----Q-----
A_TZ.TAN142         --V-----Q-----
A_RW4.538cons       --V-----Q-----
A_RW4.226           --V-----Q-----
A_RW4.074           --V-----Q-----
A_RW2.561Acons      --V-----Q-----
A_RW2.1613con       --V-----Q-----
A_RW1.W2RW009       --V-----Q-----
A_NL1.TZ94016       --V-----Q-----
A_KE1.NA114         --V-----Q-----
A_KE1.NA111         --V-----Q-----
A_KE.KEN970         --V-----Q-----
A_GA1.VI685         --V-----Q-----
A_CI2.CI42          --V-----Q-----
A_TZ.TAN9CON        R-V-----Q-----
A_UG2.72            R-V-----Q-----
A_NL1.AO9246        R-V-----Q-----
A_UG5.94UG011       --V-L---Q-----
A_BJ1.249           --------Q----I
A_RW4.730cons       --V-----Q----I
A_RW2.R2235W        T-------Q-----
B_BR1.8623          T-------Q-----
A_RW2.R2235         T-------Q-----
A_KE.KEN980         X-------Q-----
A_KE.KEN968         R-------Q----S
A_KE.KEN984         --------Q----S
D_NL.A1             R---V---Q-----
F_NL1.ZR8908        R---V---Q-----
A_ZR3.K114          R----A--Q-----
A_NL1.RW8935        Q-V-----Q-----
A_UG2.119           Q-V-----K-----
A_TZ3.1577          T-V-----Q--F--
A_RW4.081cons       --VS----Q-----
A_RW4.082cons       --VS----Q-----
A_RW3.PVP3          --VS----Q-----
A_CI2.CI2           --VN----Q-----
A_CI2.CI14          --VP----Q-----
A_KE.KEN978         --VP----Q-----
G_NG1.JV832         ---P----Q-----
A_CI1.41con         --D-----Q-----
A_KE.KEN982         --AN----Q-----
I_CONSENSUS_96      --V-----QTW---
I_CYHOcon           --V-----QTW---
A_CI1.51con         --V-----QT-F--
A_TZ.TAN15CON       --V-----QTS---
A_CM1.CA1           TG------QT----
A_UG5.94UG004       T-------QT--T-
A_RU1.SHL9          --------Q-I-TG
D_ZR1.6565.sh       --------Q-I-T-
A_RU1.MLY10         --------Q-I-TN
F_RO2.RM5302        --------Q-X-T-
B_GB4.M73767719     --------E-LFT-
D_UG3.120           --------Q-LFTI
D_UG7.44            --T-----Q-YFTS
D_UG3.79            Q-T-----Q-L-T-
D_UK1.CPHL4         Q-T-----Q-L-T-
D_UG3.12            Q-T-----Q-L-T-
D_TZ3.1604          Q-T-----Q-L-T-
D_NL2.ZR8911        Q-T-----Q-L-T-
D_ZR1.6555          Q-T-L---Q-L-T-
D_GA1.G141          Q-T-----Q-LFT-
D_NL2.AO93021       Q-T-----Q-LFT-
D_NL2.UG94015       Q-------Q-L-T-
```

D_UG3.7	Q-------Q-L-T-
D_UG3.122	Q-V-L---Q-L-T-
D_TZ1.TAN1	Q-TR----Q-L-TN
D_TZ1.TAN12	Q-TR----Q-LFTS
D_KE1.KEN966	Q-TR----QTLFT-

3

D_TZ2.005	QRT---S-Q-L-T-
D_UG4.WHO1523	QRT---T-Q-L-T-
D_CONSENSUS_96	QRT-----Q-L-T-
D_TZ2.012	RRT---T-Q-L-T-
D_UG2.2999	QRT---T-Q-LHT-
D_TZ2.030	QRT---S-Q-LFT-
U_CF1.4040	QRT---S-Q-IFT-
D_UG1.W2UG059	QRT---T-Q-Y-TK
D_UG6.9802	HRT---S-Q-L-TP
D_UG2.4133	IRT---S-Q-Y-R-
E_TH5.CMU07	TRT---G-Q-Y-R-

4

E_TH9.83	T--T----QV--R-
E_UY.UR4	T--T----QV--R-
E_TH9.179	T--T----QV--R-
E_TH9.126	T--T----QV--R-
E_TH9.115	T--T----QV--R-
E_TH5.CMU0	T--T----QV--R-
E_TH3.W2TH022	T--T----QV--R-
E_TH3.W2TH019	T--T----QV--R-
E_TH3.W2TH01	T--T----QV--R-
E_TH3.W2TH009	T--T----QV--R-
E_TH3.W2TH006	T--T----QV--R-
E_TH2.TN244	T--T----QV--R-
E_TH2.TN242	T--T----QV--R-
E_TH2.TN240	T--T----QV--R-
E_TH2.TN24	T--T----QV--R-
E_TH2.TN239	T--T----QV--R-
E_TH2.TN238	T--T----QV--R-
E_TH10.92TH011	T--T----QV--R-
E_TH1.TA8176	T--T----QV--R-
E_TH1.8673	T--T----QV--R-
E_TH1.867	T--T----QV--R-
E_TH1.8663	T--T----QV--R-
E_TH1.8659	T--T----QV--R-
E_TH.N764	T--T----QV--R-
E_MY.9214103	T--T----QV--R-
E_MY.1786	T--T----QV--R-
E_JP1.JNIH4T	T--T----QV--R-
E_JP1.JNIH2T	T--T----QV--R-
E_CONSENSUS_96	T--T----QV--R-
E_TH1.8657	T--T----QI--R-
E_TH3.W2TH018	T--T----QI--R-
E_TH3.W2TH020	T-LT----QV--R-
E_VN2.VN2	T-LT----QV--R-
E_TH3.W2TH02	T-LT----QV--R-
E_TH9.119	T-VT----QV--R-
E_TH3.W2TH003	T--TV---QV--R-
E_TH8.0289	T--TL---QV--R-
E_TH3.W2TH015	TG-T----QV--R-
E_TH3.W2TH023	T--N----QV--R-
E_UY.UR6	T--N----QV--R-
E_TH3.W2TH007	T--N----QV--R-
E_TH3.W2TH00	T--N----QV--R-
E_TH10.92TH001	T--N----QV--R-
E_TH9.97	T--A----QV--R-
E_VN2.VN1	T--A----QV--R-
E_TH8.0103	T--A----QV--R-
E_TH.T8178	T--S----QV--R-
E_TH4.D3TH966	T-TT----QV--R-
E_TH9.81	T--T----QV--K-
E_TH9.84	T--T----QV--K-
E_TH1.TA7794	T--T----QV--K-
E_TH3.W2TH024	T--T----QVL-K-
E_TH9.135	T--T----QVL-R-
E_JP1.JNIH3J	T--T----QVS-R-
E_TH5.CMU05	T--R----QV--R-
E_TH6.194	T--R----QV--R-
E_TH4.D3TH976	T--R----QV--R-
E_TH5.KH009	T-LR----QV--R-
E_TH4.D3TH975	T-VR----QV--R-
E_US.POC30506	T-VR----QV--R-
E_TH9.111	T--R----QVW-R-
E_TH9.89	T--R----QT--R-
E_NL.TH94037	T-------QV--R-
E_TH9.106	T-------QV--R-
E_TW1.286	TG------QV--R-
E_TH3.W2TH005	T--P----QV--R-
E_VN1.CT5	T--P----QV--R-
E_TH1.TA8173	T--P----QV--R-
E_TH1.8683	T--P----QV--R-
E_TH5.CMU04	T--P----QVL-R-
E_TH2.TN235	T--P----Q---R-
E_TH5.KH003	T--P----Q---R-
E_TH1.TA7792	P--T----QV--R-
E_TH2.TN243	P--TX---QV--R-
E_VN2.VN3	T--TM---QV--R-
E_VN2.VN4	T--TM---QV--R-
E_VN1.DN3	T--TM---QV--R-
E_UY.UR7	T--TM---QV--R-
E_UY.UR2	T--TM---QV--R-
E_TH5.KH012	T--TM---QV--R-
E_TH5.CMU03	T--TM---QV--R-
E_TH1.TA1067	T--TM---QV--R-
E_VN1.BX6	TX-TM---QV--R-
E_TH5.KH007	TK-TM----V--R-
E_TH6.1012	TR-T-----V--R-
E_TH7.1110	TR-T-----V--R-
E_TH1.8639	TR-T----QV--R-

E_TH6.0182	T--T-----V--R-
E_TH7.1018	T--T-----V--R-
E_TH.CM240	T--T-----V--R-
E_TH5.KH00	T--T-----R--R-
B_TH5.1948	-R-TM----VY-T-
B_US.BRVA	-R-TM----VY-T-
B_JP.ETR	-RVTM----VY-T-
B_BR6.P4	-RMTL----VY-T-
B_US18.LBV	-R-TL----VY-T-
B_GB4.M3015617	-R-SM----VY-T-
B_NL11.patNcon	-R-SL----VY-T-
B_US.BCSG3	-R-TT----VY-T-
B_AR.21281	---TL----VY-T-
B_US1.HC13	-R-TM----VL-T-
B_VE.VE6	-R-TM----VL-T-
B_TW1.358	-R-TL--R-V--T-
B_FR.CA	RR-T-----VY-T-
E_TH9.98	TR-T-----VY-K-
E_TH5.CMU08	QRVT-----VY-S-
B_NL4.BN132_84	-RMTL---KV--T-
B_US15.E	-RMT-----V--T-
E_TH.JP23A	-RMTL---HV--S-
B_HT.RF	---TK----VI---
E_TH5.KH01	IR-TR----VV-R-
B_US.SF33	RR-TS---KVL-T-
E_TH5.CMU06	TRMTM---HV--R-
E_TH5.KH014	TRMTM---HV--K-
E_TH5.KH015	TKMTM---HVV-K-
E_TH5.CMU02	I-MTR---HV--R-
E_TH5.KH004	TKMTR---HV--K-

5

B_US7.565	---P--------T-
F_AR2.16	---P--------T-
B_US4.ZhuPtA	---P--------T-
B_US11.725	---P--------T-
B_TH4.N758	---P--------T-
B_SE2.R3	---P--------T-
B_NO1.44	---P--------T-
B_NO1.40	---P--------T-
B_NO1.37	---P--------T-
B_NO1.25	---P--------T-
B_NL9.h1135	---P--------T-
B_NL7.NE1058con	---P--------T-
B_NL5.pt6con	---P--------T-
B_NL4.I7025_9	---P--------T-
B_NL4.I4008_86	---P--------T-
B_NL4.I1057_87	---P--------T-
B_NL4.H495_85	---P--------T-
B_NL4.H188_90	---P--------T-
B_NL4.H164_90	---P--------T-
B_NL4.H1091_90	---P--------T-
B_NL2.Wolfsh495	---P--------T-
B_NL10.GQ93013	---P--------T-
B_LT.LIT21A	---P--------T-
B_IT1.204	---P--------T-
B_IT1.190	---P--------T-
B_ES.15	---P--------T-
B_ES.09	---P--------T-
B_ES.06	---P--------T-
B_DE1.D	---P--------T-
B_BR5.510	---P--------T-
B_TW1.394	R--P--------T-
B_US4.ZhuPtR	R--P--------T-
B_NL10.GF921953	--VP--------T-
B_IT.Sala2	---P-----Q--T-
B_NL5.pt5con	---P------L-T-
B_ES.S1	---P--------TI
B_IT1.145	---P--------TG
B_GM.GM6	---P-------FT-
B_HT2.H6018	---P-------FT-
B_AU1.MRC2	---P-------WT-
B_US24.85W6A	---P----------
B_VN1.HCM9	---P----------
B_US22.E0105	---P----------
B_US2.D2US725	---P----------
B_US14.3	---P----------
B_US13.141con	---P----------
B_US1.HC44	---P----------
B_US1.HC24	---P----------
B_SE2.R4	---P----------
B_RU2.2236	---P----------
B_NO1.29	---P----------
B_NL4.H186_85	---P----------
B_NL2.Wolfsh	---P----------
B_LT.LIT18A	---P----------
B_HT2.H6012	---P----------
B_HT2.H6010	---P----------
B_HT1.D1HA652	---P----------
B_GB4.M2347015	---P----------
B_ES.10	---P----------
B_ES.07	---P----------
B_US22.E0803	---P-----V----
B_RU2.4035	--VP----------
B_US18.LBLCT	--VP----------
A_RW1.W2RW016	--VP----------
B_US.MA	---PM---------
B_US.TN1006	---PM---------
B_CY1.HO464	---PM---------
F_BR.RJI03	---PL---------
F_NL1.BR94009	---PL---------
B_NL11.patFcon	---PL---------
B_GB.V9	---P-------I--
B_US11.552	---P-------F--
B_NL4.I1025_89	R--P----------
B_US15.F	R--P----------
B_US22.E0903	R--P----K-----
B_FR.CD	---P----K-----

```
B_US1.HC09          Q--P----K-----
B_CH1.K6            ---P----K---T-
B_NL1.A7            R--P----K--FT-
B_TH6.0286          ---P----AF-RTG
B_BR3.HRJ636        ---P-A--S-WF--
B_BR3.RJ636         ---P-A--S-WF--
B_US15.D            ---PM-------T-
B_US2.D2US665       ---PM-------T-
B_US1.HC08          ---PM-------T-
B_NL4.I4004_88      ---PM-------T-
B_NL10.SR9235       ---PM-------T-
B_GB4.JB14          -G-PM-------T-
B_HT2.H5986         ---PM---K---T-
B_US1.HC05          R--PM---K---T-
B_NL10.SR9115       R--PM----V----
B_CH1.K32           ---P----KVL-T-
B_DE1.J             ---P---EKV--T-
B_NL4.H1112_90      R--P--L-S---T-
B_US1.HC19          -G-P--------T-
B_US1.HC40          -G-P--------T-
B_TH4.N752          -G-P--------T-
B_ES.III            -G-P----------
B_TH4.N753          -G-P----------
B_DE2.3497          XG-P--------XX
A_RW3.PVP4          ---P----G-----
B_BR2.W2BR004       -G-P----GS----
```

6

```
F_RO2.RM5306        ----L---Q-----
F_RO2.RM5307        ----L---Q-----
F_RO2.RM53035       ----L---Q-----
F_RO2.RM53034       ----L---Q-----
F_RO2.RM53032       ----L---Q-----
F_RO2.RM5303        ----L---Q-----
F_RO2.RM53029       ----L---Q-----
F_RO2.RM53027       ----L---Q-----
F_RO2.RM53022       ----L---Q-----
F_RO2.RM53018       ----L---Q-----
F_RO2.RM53013       ----L---Q-----
F_RO2.RM5301        ----L---Q-----
F_RO2.RM53002       ----L---Q-----
F_RO1.14046         ----L---Q-----
F_RO1.1404          ----L---Q-----
F_RO1.14036         ----L---Q-----
F_RO1.14028         ----L---Q-----
F_RO1.14027         ----L---Q-----
F_RO1.14024         ----L---Q-----
F_RO1.14020         ----L---Q-----
F_RO1.14018         ----L---Q-----
F_CONSENSUS_96      ----L---Q-----
F_RO2.RM53040       --------Q-----
F_RO2.RM53043       --------Q-----
F_RO2.RM53037       --------Q-----
F_RO2.RM53015       --------Q-----
A_UG5.94UG016       --------Q-----
A_TZ3.1584          --------Q-----
A_KE.KEN3           --------Q-----
A_KE.KEN112         --------Q-----
A_IN1.CMCH14        --------Q-----
A_GA1.G41           --------Q-----
A_CF1.4010          --------Q-----
B_US11.419          --------Q-L---
G_GA.LBV217         --------Q-L---
G_BJ1.43            ----F---Q-L---
G_GB1.22            ----F---Q-----
B_TH1.864           ----L---Q-W---
B_TH7.156           ----L---Q-W---
B_TH7.134           --------Q-W---
A_NL1.GH95012       A---M---Q----P
F_RO2.RM53014       ---XL---Q---T-
G_RU.RUS12A         ---NL---Q---T-
G_GB.K              ---NL---Q-I---
B_NL5.pt7con        ---PM---K-----
B_TW1.384           ---PM---K-----
B_HT2.H6004         ---PM---K-----
B_AR1.33            ---PM---K-----
B_US2.D2US660       ---PM---K-M---
B_CH1.K72           ---SM---K-I-T-
B_US1.HC07          ---TM---KV---G
B_US18.LBF          ----M---KV---G
B_US4.ZhuPtV        ---TM---KV--V-
G_BJ1.259           ---TF---Q-----
G_CONSENSUS_96      ---TF---Q-----
G_CF.4067           ---SF---Q-----
G_NL1.GH94012       -R-SF---Q-----
H_CONSENSUS_96      ---S----Q-----
H_ZR.VI557          ---S----Q-----
C_CF1.15166         ---SL---Q-----
G_NG1.G9            ---TL---Q-----
G_NL1.ZR9119        R--TL---Q-----
G_NG1.JP882         ---TFA--Q-----
G_RU.131con         ---TFA--Q-L---
G_RU1.BUK3a         ---TF---Q-I---
B_ES.VI             ---AL---G-V---
```

7

```
B_US5.pt6bl         ---N----------
U_BR3.RJI0          ---N----------
B_US22.E1001        ---N----------
B_US2.D2US705       ---N----------
B_NL9.h1081         ---N----------
B_NL5.pt9con        ---N----------
B_NL5.pt3con        ---N----------
B_NL4.H748_90       ---N----------
B_HT2.H5998         ---N----------
B_HT2.H5990         ---N----------
B_DE2.3498          ---N----------
B_CY1.HO294         ---N----------
```

B_BR3.HRJI01	---N----------
B_US12.CHB2A	---N------W---
B_US24.8674A	---N------W---
B_TH7.184	---N------W---
B_HT1.D1HA65	---N------W---
B_CY1.HO27	---N---------I
B_US22.E0204	---N-------V--
B_US.SF162	---T----------
B_US6.I02	---T----------
B_BR2.W2BR026	---T----------
B_CH1.K28	---S----------
B_NL3.NET3	---S----------
B_BR1.10593	---S----------
B_NL4.I5037_89	---X----------
B_NO1.24	---Q----------
A_ZR.Z32	---S--------F--
B_US12.LC2A	R--N----------
B_US6.I03	R--N----------
B_US.TN1009	R--N----------
B_US.Diaz	R--N----------
B_SE1.295	R--N----------
B_DE.serocons	R--N----------
B_CY1.HO503	R--N----------
B_NL5.pt17con	R--NM---------
B_US1.HC28	R--T----------
B_US22.E11A06	R--T----------
B_NO1.31	R--T----------
B_CY1.HO042	R--T----------
B_IT1.130	R--S----------
B_BR2.W2BR014	R--I----------
B_NL3.NET2	R--G------I---
F_AR1.20016	---QX-------X-
B_US1.HC22	---T--------T-
B_US24.87Y3E	---T--------T-
B_SE2.R5	---T--------T-
B_SE1.2815	---T--------T-
B_NL9.h571	---T--------T-
B_NL4.I221_90	---T--------T-
B_ES.18	---T--------T-
B_US.JFL	---TL-------T-
B_BR1.10575	--LSL-------T-
B_TW1.646	-R-IL----S--T-
B_CH1.P6	---TM----T----
B_GB5.4660	---T-----T--TG
B_US21.P1	---Q--------T-
F_BR3.93BR029	---Q--------T-
B_NL4.B63_84	---Q--------T-
B_BR1.10553	---Q--------T-
F_BR.7944	---QL-------T-
F_AR2.18	---Q----Q---T-
A_KE.KEN973	---R----Q---T-
A_KE1.NA118	---R----Q---TS
B_US18.LBZ	---R------W-T-
U_CONSENSUS_96	---R--------T-

A_CY.HOcon	---RF---Q---TN
U_ZM.ZAM184	R--RF---Q---TN
U_KE.KEN976	---RF---Q---TS
F_AR2.15	R--RF---Q-----
G_KP.Kr121	---RF---Q-----
B_NO1.O5	---RF---S---T-
B_BR4.BZ167A	RR-R-----T--TG
U_CF1.4087	R-MR-----T--T1
B_ZA.ZA504	-VYR--------TI
B_CN1.062	---Y-------HT-
B_US.SF2	---Y-------HT-
B_CN1.047	---Y-------HT-
B_CN1.046	---Y-------HT-
B_CN1.033	---Y-------HT-
B_CN1.032	---Y-------HT-
B_CN1.030	---Y-------HT-
B_CN1.021	---Y-------HT-
B_CN1.019	---Y-------HT-
B_CN1.014	---Y-------HT-
B_CN1.007	---Y-------HT-
B_CN1.166	---YL------HT-
B_CN1.196	---YL------HT-
B_CN1.045	---Y-------NT-
B_CN1.071	---Y-------NT-
B_CN1.042	---YL------NT-
B_CN1.072	---YL-------T-
B_CY1.HO25	-G-YM------RT-
B_CN1.073	-C-YL------NT-
B_ZA.ZA513	RC-Y-------HT1
B_CN1.016	---Y------SHT-
B_CN1.059	---Y------SHT-
B_CN1.015	---Y------SHT-
B_CN1.080	---Y------CHT-
B_PY.1284	---F-------HTA
F_BR1.BZ126	---YF------HTA
B_GB.Man	---Y-----R-HV-
B_ES.P9	----------HT-
B_US2.D2US704	----------HT-
B_PY.3615	RR---------HT-
B_US.ALA	GR---------HT-
B_GA.OYI	NR-S-------HT-
B_HT1.D2HA593	-R-S-------R--
B_HT2.H13958	-R-S-------R--
B_US12.CHBMOM	-R-S-------RT-
B_VE.VE8	-R-S-------RTK
B_ES.D17	-RVS------WRT-
B_ES.IX	-RVS-------RTA
B_HT1.D2HA594	-R-S-----VW-T-
B_HT1.D2HA596	-R-S-----VW-T-
B_HT2.H13962	-R-S------W-T-
B_HT2.H13960	-R-S------S-T-
B_BR3.SPB4	-RKS--------T-
B_US1.HC06	-RLS-------V--
B_US20.C3con	RR-S-------V--

```
B_US.SBB            -R-S--------V-A
B_US.P896           RRLS---------R
B_US1.HC3           RR-S--------T-R
B_GB.V87            ER-S--------I-R
B_GB4.WB29          R--T--------H--
B_HT2.H5992         R--T--------D-Y
B_US.WMJ22          R-LS--------RTR
B_US2.D2US724       R--SM-------RTG
B_DE1.C             RR-------------
B_ES.P14            RR-------------
B_JP.JH32           RR-----------T-
B_US.WR27           RR-----------TD
A_UG.UG06           RR--------S--TS
B_KP.Kr111          -R----------VT-
B_TW1.138           RR--------S-VTG
A_CF2.SAS           RRM---------I--
B_BR6.P7            RR----------IG-
B_GB4.M2686401      R-F---------LTE
B_FR.LAI            IR-QR-------VTI
A_NL1.UG9283        RR-P-----Q--H-R
B_JP.GUNA           ---T--S----H-I
```

8

```
E_CF.90CR402        I-AR------V-HT-
E_CF1.4002          I-AR------V-HT-
E_CF1.4069          I-AR------V-HK-
E_CF2.4071TG2       T-VR------V--K-
E_TH5.KH005         I-TR------V--K-
E_CF1.4013          T-VR------V-HK-
E_TH5.E1058         T-TR------V-HK-
E_TH5.KH008         T-MR------V-HR-
E_UK1.11643         T-FR----QV-HK-
B_US19.VA2          R-VR----G-W-K-
B_US23.201          R-VR----G-MFR-
```

9

```
B_CU.95CU043        -N-YL-R--SV---
B_ZA.ZA512          -G-Y--Q---V-T-
B_TH2.TB132         -X-XM----XX-T-
B_BR.002con         ---RM-L--WSVYA
B_CY1.HO11          -R-RF-L--FVTTG
```

10

```
O_FR.CFBCF03        -Q-G---MSVYSGS
```

11

```
B_ZA.ZA508          ------------T-
F_BR2.HSP209        ------------T-
B_US9.S4            ------------T-
B_US9.S             ------------T-
B_US8.R             ------------T-
B_US7.657           ------------T-
B_US6.I05           ------------T-
B_US3.D2US716       ------------T-
B_US3.D2US715       ------------T-
B_US3.D1US712       ------------T-
B_US24.85W3F        ------------T-
B_US22.E0607        ------------T-
B_US22.E0502        ------------T-
B_US21.P6           ------------T-
B_US21.P5           ------------T-
B_US21.P4           ------------T-
B_US21.P3           ------------T-
B_US2.D2US662       ------------T-
B_US18.LBQ          ------------T-
B_US18.LBPTCC1      ------------T-
B_US18.LBK          ------------T-
B_US18.LBJ          ------------T-
B_US17.CB8          ------------T-
B_US15.C            ------------T-
B_US14.2            ------------T-
B_US11.550          ------------T-
B_US11.333          ------------T-
B_US1.HC45          ------------T-
B_US1.HC35          ------------T-
B_US1.HC33          ------------T-
B_US.TN1007         ------------T-
B_US.TN1003         ------------T-
B_US.JRCSF          ------------T-
B_US.BAL            ------------T-
B_US.ADA            ------------T-
B_TW1.237           ------------T-
B_TH6.1615          ------------T-
B_TH5.018           ------------T-
B_SK1.BTS18         ------------T-
B_SE2.R1            ------------T-
B_SE1.1433          ------------T-
B_PR1.D2PR732       ------------T-
B_PR1.D2PR729       ------------T-
B_NO1.42            ------------T-
B_NO1.36            ------------T-
B_NO1.34            ------------T-
B_NO1.30            ------------T-
B_NO1.27            ------------T-
B_NO1.22            ------------T-
B_NO1.21            ------------T-
B_NO1.12            ------------T-
B_NL9.h709          ------------T-
B_NL9.h569          ------------T-
B_NL9.h526          ------------T-
B_NL9.h424          ------------T-
B_NL9.h412          ------------T-
B_NL9.h411          ------------T-
B_NL9.h159          ------------T-
B_NL9.h1140         ------------T-
B_NL9.h1110         ------------T-
B_NL4.K999_86       ------------T-
```

```
B_NL4.I3048_89      ----------T-
B_NL4.I3007_86      ----------T-
B_NL4.I257_9        ----------T-
B_NL4.I1164_90      ----------T-
B_NL4.H724_8        ----------T-
B_NL4.H557_82       ----------T-
B_NL4.H466_82       ----------T-
B_NL4.H457_85       ----------T-
B_NL4.H36_82        ----------T-
B_NL4.H35_80        ----------T-
B_NL4.H339_82       ----------T-
B_NL4.H293_85       ----------T-
B_NL4.H153_82       ----------T-
B_NL4.BN93_85       ----------T-
B_NL4.BN90_85       ----------T-
B_NL4.BN152_84      ----------T-
B_NL4.BN142_84      ----------T-
B_NL4.BA206_84      ----------T-
B_NL4.BA1_84        ----------T-
B_NL3.NET           ----------T-
B_NL12.0337         ----------T-
B_NL12.0230         ----------T-
B_NL12.0211         ----------T-
B_NL10.SR9269       ----------T-
B_NL10.MA9136       ----------T-
B_NL1.A5            ----------T-
B_NL.wolfscon       ----------T-
B_NL.h159           ----------T-
B_LT.LIT17A         ----------T-
B_HT2.H6014         ----------T-
B_HT2.H6006         ----------T-
B_HT2.H6002         ----------T-
B_GB2.CPHL2         ----------T-
B_ES.P6             ----------T-
B_CY1.HO28          ----------T-
B_CONSENSUS_96      ----------T-
B_CH1.P             ----------T-
B_CH1.K1            ----------T-
B_BR3.RJ54          ----------T-
B_BR3.RJ379         ----------T-
B_SK1.BTS23         R----------T-
B_US1.HC43          R----------T-
B_NL4.I1083_87      R----------T-
B_NL3.NET6          R----------T-
B_IT1.115           R----------T-
F_BR2.HSP255        ----L-------T-
F_BR2.HSP255A       ----L-------T-
B_HT2.H6000         ---------S--T-
B_HT2.H13954        ----------TG
B_TT.QZ4589         ----------TG
B_HT1.D2HA590       ----------TG
B_ES.P21            ----------TS
B_NL1.A10           --------S---T-
B_US1.HC36          --------S---T-

B_ES.THF17          --------S---T-
B_US2.D2US727       --V---------T-
D_UG3.109           Q-V---------T-
U_AR.20021          XX----------T-
B_NL5.pt10con       ----------FT-
B_US18.LBM          ----------FT-
B_HT2.H5988         ---------T-FT-
B_IT1.20            XG---------FT-
B_US9.S3            ----------
F_CM.CA20           ----------
B_US23.303          ----------
B_US22.E04B03       ----------
B_US2.D2US664       ----------
B_US2.D2US656       ----------
B_US18.LBG          ----------
B_US17.CB9          ----------
B_US14.4            ----------
B_US10.SFPEcon      ----------
B_US1.HC39          ----------
B_US1.HC34          ----------
B_US1.HC26          ----------
B_US.twinABcon      ----------
B_US.TN1002         ----------
B_TW1.382           ----------
B_SE1.93            ----------
B_RU2.4439          ----------
B_RU2.190           ----------
B_PY.12845          ----------
B_PR1.D2PR733       ----------
B_PR1.D2PR728       ----------
B_NO1.O9            ----------
B_NO1.O6            ----------
B_NO1.45            ----------
B_NO1.26            ----------
B_NO1.16            ----------
B_NL9.h1145         ----------
B_NL8.594           ----------
B_NL4.I3064_90      ----------
B_NL4.I3008_88      ----------
B_NL4.I1020_86      ----------
B_NL4.H594_85       ----------
B_NL4.H552_90       ----------
B_NL4.H17_82        ----------
B_NL4.H169_85       ----------
B_NL4.BN18_84       ----------
B_NL4.BN130_85      ----------
B_NL10.UM94038      ----------
B_NL10.UM94030      ----------
B_IT1.199           ----------
B_IN.IND9           ----------
B_ID.1701           ----------
B_HT2.H6024         ----------
B_GB5.4663          ----------
B_GB5.4659          ----------
```

```
B_GB5.4656          --------------
B_FR.J91            --------------
B_FR.J6             --------------
B_ES.106            --------------
B_CH1.P5            --------------
B_CH1.P3            --------------
B_CH1.K53           --------------
B_CH1.K26           --------------
B_BR3.HSP205        --------------
B_BR2.W2BR020       --------------
B_BR2.W2BR017       --------------
B_RU2.4466          --V-----------
B_RU2.4714          --V-----------
B_NL4.I3077_9       -X------------
B_CH1.K9            -------------S
B_CY1.HO39          -----------F--
B_RU2.10120         --------K-----
B_RU2.5852          --------K-----
B_RU.RUS3A          --------K-----
B_NL9.H239          --------K-----
B_NL4.H1_85         --------K-----
B_NL1.A             --------K-----
B_LT.LIT11A         --------K-----
B_BR2.W2BR023       --------K-----
A_UG.1033           ----L---------
B_US3.D2US71        ----L---------
U_CF1.4056          T---L---------
B_US1.HC20          -----A--------
B_US1.HC4           -----A--------
B_US1.HC16          -----A--------
B_US.TN1005         -----A--------
B_TH7.91            -----A--------
B_SE1.930           -----A--------
B_PY.3616           -----A--------
B_NL4.I5032_87      -----A--------
B_DE2.3493          -----A--------
B_CH1.K12           -----A--------
B_US1.HC30          R----A--------
A_CF1.4044          T----A--------
U_SE.KI4803         RIM---------K-
B_US.JM             -----A------T-
B_US18.LBE          -----E------T-
B_SE.pt11s113       ---N-R------T-
B_CY1.HO21          ---N-E--------
B_US18.LBLCN        -----E--------
B_US13.149con       R---LQ------SS
B_US22.E03A03       --------S-----
B_ZA.ZA524          --------S-----
B_BR2.W2BR018       --------S-----
B_BR1.8625          --------S-----
B_RU2.3818          -----A--S-----
B_RU2.4261SE        -----A--GT----
B_US1.HC37          ----LA---G-H--
B_US2.D2US726       R----A-----H--
B_GB.V77            ---------V-H--
B_US1.HC25          --------K--D--
B_US12.LC1A         ----------D--
B_SK1.BTS9          --V-L--K-V-FT-
B_PY.12847          ----M---------
B_VE.VE             ----M---------
B_CZ.BTSPR          ----M---------
B_US2.D2US714       R---M---------
A_CF1.11699         --V-M---------
B_US11.UK           ----M---K-Y---
A_CF1.4018          T-V-M---KT----
B_CI.CI22           ----M---S-I---
B_HT2.H6016         ----M---S-----
B_TW1.257           --V-M---STL---
B_ES.P20            ----M----TVH-I
A_ZR1.6569          RG--M---QIL---
J_SE1.7887          -G--M---QVL---
J_SE1.7022          E---M---QVL---
B_PY.3614p          TR-QM----VL---
```

12

```
B_GB.CAM            ---A-----TV---
B_US1.HC38          -G-A-----TV---
B_US.NY5CG          -G-A-----TL--R
B_US.ACP            -G-G-----TV-TA
B_ZA.ZA509          -G-G-----TV-TA
A_UG2.84            ---G----QT---A
C_BU1.91BU007       -G-G----QT-F--
F_AR1.21280         XX-XX----X--X-
```

13

```
B_US20.C1con        ---N--------T-
B_US8.R2            ---N--------T-
B_US2.D2US706       ---N--------T-
B_US18.LBAcon       ---N--------T-
B_US17.CB5          ---N--------T-
B_US11.074          ---N--------T-
B_NO1.41            ---N--------T-
B_NO1.38            ---N--------T-
B_NL9.h57           ---N--------T-
B_NL4.I5038_89      ---N--------T-
B_NL4.I1035_87      ---N--------T-
B_NL4.H29_82        ---N--------T-
B_NL12.0617         ---N--------T-
B_NL12.0157         ---N--------T-
B_NL10.SR925752     ---N--------T-
B_NL1.A3            ---N--------T-
B_IT2.17B           ---N--------T-
B_GB5.4666          ---N--------T-
B_GB5.4665          ---N--------T-
B_CH1.P7            ---N--------T-
B_CH1.P2            ---N--------T-
B_BR1.7942          ---N--------T-
```

B_NO1.O4	---N-----V--T-
B_NL4.I3072_90	---N-------FT-
B_US17.CB	---S--------T-
B_US18.LBSDIR	---S--------T-
B_US.TN1000	---S--------T-
B_US.SBC	---S--------T-
B_TW1.391	---S--------T-
B_SE2.R2	---S--------T-
B_NL12.0008	---X--------T-
B_NL12.0583	X--N--------T-
B_US.SF128	---N------I-T-
B_US.YU	---N------L-T-
B_NL4.I138_89	---NT-------T-
B_US17.CB6	R--N--------T-
B_US2.D2US717	R--N--------T-
B_US.SBA	R--N--------T-
B_SK1.BTS28	R--N--------T-
B_IT2.32B	R--N--------T-
B_IT2.13A	R--N--------T-
B_NL7.NE537con	R--N-------WT-
B_BR1.8615	R--T--------T-
B_ES.03	R--T--------T-
B_NL9.h39	R--S--------T-
B_US24.88B4E	----------W-T-
B_US9.S2	----------W-T-
B_US1.HC12	R---------W-T-
B_TH7.125	---N------W-T-
B_US.TN100	R--N------W-T-
B_TH7.138	---N----Q-W-T-
F_RO2.RM53024	----L---Q---T-
F_RO2.RM5308	----L---Q---T-
F_RO2.RM53023	----L---Q---T-
F_RO1.14034	----L---Q---T-
F_BR2.HSP238	----L---Q---T-
F_BR2.HSP229	----L---Q---T-
B_CN1.149	----L---Q---T-
B_TW1.386	--------K---T-
B_US7.306	--------K---T-
B_TW1.366	--------K---T-
B_NL5.pt1con	--------K---T-
B_HT2.H5994	----M---K---T-
B_NL8.672	----M---K---T-
B_NL9.h716	----L---K---T-
B_IT.Sala	----M---KT--T-
B_NL4.H135_90	----M---KT--T-
B_US1.HC2	R--NM-------T-
B_US15.B	R--NM-------T-
B_US1.HC1	R--NM-------T-
B_US15.A	---NM-------T-
B_AR1.34	----M-------T-
B_BR3.HRJ17	----M-----Y-T-
B_CH1.K8	R---M---K-I-T-
B_US19.VA1con	R---M---Q-L-T-
B_US21.P2	-G----------T-
B_US4.ZhuPtL	-G----------T-
B_US2.D2US657	-G----------T-
B_US18.LBPATH2	-G----------T-
B_US16.NYRT3	-G----------T-
B_TW1.91	-G----------T-
B_TH4.N760	-G----------T-
B_TH4.N759	-G----------T-
B_TH4.N757	-G----------T-
B_TH4.N756	-G----------T-
B_TH4.N755	-G----------T-
B_TH4.N754	-G----------T-
B_SE1.1866	-G----------T-
B_NO1.O3	-G----------T-
B_NO1.43	-G----------T-
B_NO1.35	-G----------T-
B_NL4.I5020_89	-G----------T-
B_NL4.I3084_90	-G----------T-
B_NL4.BN29_85	-G----------T-
B_NL4.BA65_85	-G----------T-
B_ES.X	-G----------T-
B_DE1.F	-G----------T-
B_CY1.HO48	-G----------T-
B_CM.CA5	-G----------T-
B_AR1.13	-G----------T-
B_NL4.I33_90	RG----------T-
B_US1.HC23	RG----------T-
B_NL4.H43_90	RG----------T-
B_NO1.20	-G--M-------T-
B_US1.HC27	-G--M-------T-
B_NL4.I141_9	-G-Q--------T-
B_NL5.pt4con	-G-N--------T-
B_VE.VE5	-G-Y--------T-
B_TH6.1611	-G------K---T-
B_US.WM	-G------K---T-
B_NL4.H23_82	-G------K---T-
A_BJ1.281	-GV-----K---T-
B_BR5.504	-G------K---TR
B_DE1.E	-G------X---T-
B_ES.108	-G------N---T-
A_BJ1.241	-GV-----Q---T-
B_US11.306	RG------Q---T-
B_US.BWB	-G-NV-----L-T-
B_US12.LC3A	-G-N------V-T-
B_DE.HAN	-G--------V-T-
B_NL6.16	-G--------V-T-
B_BR6.P9	-G--------V-T-
B_NL3.NET5	RG--------V-T-
B_NL4.I3057_9	RG--------V-T-
B_ES.IV	-G--------L-T-
B_VE.VE2	QG--------I-T-
B_NO1.11	-G--L---K-I-T-
B_US13.144con	-G--L-----I-T-
B_MY1.MRNP05	-G--L---Q-L-T-
D_ZR.MAL	RG--F---Q-L-T-

```
D_ZR1.6651.sh       -G------Q-L-TS
B_NL.168            -R----------T-
B_US.MN             -R----------T-
B_NL5.pt8con        -R--------V-T-
B_TH7.140           -R------Q-L-T-
B_BR1.7940          -R-Y------V-T-
B_GB.V82            -R-Y------V-T-
B_US5.pt5bl         -R-A------V-T-
B_NL4.H490_85       ----L---K-V-T-
B_US11.44           ----L----SV-T-
D_BI.BU009con       RGT---L---Y-T-
U_BU1.91BU009       RGT---L---Y-T-
B_NL1.A4            RG----L--R--TR
B_VE.VE7            RG--L-L--R--T-
D_TZ1.TAN6          QGT-------Y-T-
D_UG8.94UG013       QGT-------Y-T-
D_TZ1.TAN3          QGT-------Y-T-
D_TZ1.TAN2          QGT-------Y-T-
D_KE1.KEN97         QGT-------Y-T-
D_KE.NA116          QGT-------YFT-
D_UG3.110           QGT-------YWT-
D_TZ1.TAN7          Q-T-F-----Y-T-
D_TZ1.TAN           QGA-------Y---
D_CI.CI13           QGT-----Q-L-T-
D_NL2.ZR9291        QGT-----Q-L-T-
D_KE1.KEN986        QGT-----QTLFT-
D_UG1.W2UG035       EGT-------LFT-
U_UG1.92UG0352      EGT-------LFT-
D_UG3.74            QGT-M-----L-TI
D_UG8.94UG010       Q-T-M-----L-T-
```

14

```
B_US1.HC29          R-------------
B_US22.E12A03       R-------------
B_US.SC             R-------------
B_NL9.h1082         R-------------
B_NL5.pt2con        R-------------
B_NL5.pt16con       R-------------
B_HT2.H6020         R-------------
B_CY1.HO433         R-------------
B_AU1.MRC3          R-------------
B_IN1.30008         R----------F--
B_NL5.pt14con       R------------R
B_US18.LBD          R------------A
B_BR3.HRJ27.sh      R-----------GG
B_TW1.261           RH-------T---R
B_RU2.20637         R-------K---TG
B_RU2.5014          R-------K---TG
B_IT1.136           R-------K---TG
B_ES.17             R-------KX--TG
B_ES.14             R-------KV---G
B_NO1.15            RG------KV---G
A_GA1.LBV23         RE------Q----G
B_SK1.BTS20         R---L---K-----
B_US1.HC32          R-------K-----
B_US12.LC4A         R---M---K---G-
B_ES.P13            R---L---S-----
B_NL5.pt19con       R-------S-----
B_NL4.H434_85       --------G-----
B_NO1.O8            ----L---G-----
B_BR3.RJ623         R---L---G---T-
B_NL10.SR93047      R---F---AR----
B_US1.HC10          R-V-----S-L-T-
B_US20.C2con        R-------S-L-T-
B_GB5.4662          R-------S-V-T-
B_CH1.K65           --------K-L-T-
B_TH.JP23B          ----V-----L-T-
B_CH1.K23           R-V-V---K-L-T-
A_CM1.CA6           R---F---QTL---
B_US18.LBO1         R---L----TL-T-
B_GB.AIT            --------KTL---
B_ES.13             R--T----SV--TG
B_NL10.SR9038       R--P----KV--TG
B_US6.I07           ----L---SV--TG
```

15

```
B_US18.LBW          -G------------
F_BR1.BZ163         -G------------
B_US17.CB4          -G------------
B_US1.HC46          -G------------
B_US1.HC42          -G------------
B_US1.HC18          -G------------
B_US1.HC0           -G------------
B_SE1.1032          -G------------
B_RU2.4112          -G------------
B_NO1.O1            -G------------
B_NO1.39            -G------------
B_NO1.32            -G------------
B_NL9.h1054         -G------------
B_NL5.pt13con       -G------------
B_NL4.I3091_9       -G------------
B_NL4.I142_87       -G------------
B_NL4.H450_85       -G------------
B_NL4.H228_90       -G------------
B_NL3.NET4          -G------------
B_IT1.193           -G------------
B_GB5.4664          -G------------
B_FR.CB             -G------------
B_ES.V              -G------------
B_ES.26             -G------------
B_DE2.3499          -G------------
B_DE1.I             -G------------
B_BR2.W2BR030       -G------------
B_BR1.8633          -G------------
A_GA.VI19           -G------------
B_ES.P19            -G-------V----
F_NL1.ZR93069       -G-------V----
B_BR5.515           -G-------V----
```

```
B_CY1.HO45      -G-------T----
B_ES.20         -G-------T----
B_CH1.K47       -G-------T----
B_BR5.513       -G-------T----
B_BR1.1058      -G-------T-F--
B_US18.LBY      -G---A---T----
B_GB.V74        RG-----------
B_NL5.pt11con   RG-----------
B_ES.THF13A     RG----------G-
B_GB5.4658      XG-----------
B_NL4.I117_90   IG-----------
B_US11.107      RG-N---------
B_US6.I04       RG-N---------
B_ES.P17        RG-N---------
B_NO1.10        -G-N---------
B_US17.CB2      -G-N---------
B_HT2.H6008     -G-T---------
B_NL4.I90_89    -G------S-----
B_US18.LBB      -G------S-----
A_CM1.CA        -G------S-I---
B_US23.811      -G--------I---
A_UG5.94UG003   -GV-----Q-----
A_UG5.94UG018   -GV-----Q-----
A_UG2.118       -GV-----Q-----
A_NG.IBNG       -GV-----Q-----
A_BJ1.251       -GV-----Q-----
A_BJ1.260       -GV-----QT----
A_CI1.39con     -G------Q-----
A_RU1.GAN1      -G------Q-----
A_KE.KEN983     -G------QS----
J_GM1.GM5       -G------QVL---
J_GM1.GM7       -G------QVL---
J_GM1.GM4       -G------QVL---
J_CONSENSUS_96  -G------QVL---
A_BJ1.233       -G------Q----A
A_RW4.566cons   -G------Q----A
A_GA1.G135      RG------Q-----
A_ZR3.P104      RG--L---Q-----
B_ES.P18        -G------K-----
B_NL4.I70_88    -G------K-----
B_NL6.320       -G--M---K-----
B_NL5.pt15con   -G------AFYATG
B_SK1.BTS11     -G------KFYATG
B_SK1.BTS12     -G--M---KF-ATG
B_US16.NYRT     -G-----WGT-F--
B_HT2.H6022     -G------G---T-
B_RU2.999       RG------G---T-
B_US18.LBBB     -G------G--FT-
B_BR1.1059      -G--M---G---T-
B_ES.P12        -G--L---G---T-
B_HT2.H5996     -G------G-----
B_US6.I0        -G--L---G-----
B_ES.VIII       -G--L---G--F--
B_NL1.A2        -G--L---GT----
B_RU2.154       ----M---GT----
B_BR1.1056      -G--M---ST----
B_RU.RUS4A      ----M---G-----
B_US17.CB3      ----M---G-L---
B_US22.E0704    -G--M---G-M---
A_ZR3.PZ61      -G--L---QT-C--
B_TH7.114       -G--L---Q-LC--
B_AR1.23        -G--FA---TP---
```

16

```
B_DE1.A         -D--------I---
B_FR.CC         -D------------
B_US14.         ----------I---
F_BR1.BZ162     ----------L---
B_AU1.MRC       -D--------IFR-
B_PY.12842      -G------K-Y-T-
D_UG1.W2UG005   -G--------Y-T-
D_UG3.114       QG--------Y-TD
D_NL2.UG93071   -G------Q-YFTA
F_GA.VI354      -G-R-----VI---
F_GA1.VI354     -G-R-----VI---
B_VE.VE3        -G-G-----V----
B_JP.KM03       -G-R------V--A
B_NL4.H187_90   -G-R-------I-A
B_ES.R1         RGV----R-G--T-
B_PY.12838      -G-R---R-GTVY-
```

17

```
B_BR3.HSP228A2  ----M-W-------
B_BR3.SPB       ----M-W-------
B_BR2.W2BR02    ----M-W-------
B_BR2.W2BR003   ----M-W-------
B_BR1.8635      ----M-W-------
B_BR1.7936      ----M-W-------
B_BR1.7932      ----M-W-------
B_BR.BZ         ----M-W-------
B_BR2.W2BR019   R---M-W-------
B_BR3.RJ49      ----M-W-----T-
B_BR3.SPB2      ----M-W-----T-
B_BR5.512       ----M-W-----TN
B_BR3.RJ59      ----L-W-------
B_BR3.RJ62      ----L-W-------
B_BR2.W2BR024   ----L-W-------
B_BR1.862       ----L-W-------
B_BR3.HRJ70     ----L-W-----T-
B_BR3.RJ64      ----L-W-----T-
B_BR3.RJ478     ----M-W----HTN
B_BR5.514       ----M-W--T----
B_BR6.P3        ----M-W--T----
B_BR2.W2BR028   ----M-W--T----
B_BR1.8629      ----M-W--T----
B_BR3.SPB3      -G--M-W--T----
B_BR5.507       -G--M-W--T----
```

B_BR1.7930	-G--M-W--T----
B_AR1.35	----MAW--T----
B_AR1.29	-G--M-W-------
B_BR3.RJ484	-G----W-------
B_PY.12839	-G--L-W--T-FT-
B_BR1.7934	----V-W--SLFT-
B_BR3.RJ14	----V-W---L-T-
B_BR5.505	R---T-W---LFT-
B_BR4.BZ200A	----L-F---L---
B_BR6.P8	----V-F---L---
B_BR1.10565	-G--L-F---L---
B_AR1.24	----L-F--TL---
B_BR3.HRJI04	-G----F--TLF--
B_TW1.236	R---M-L-----T-
B_US4.ZhuPtF	R---M-L-----T-
B_BR3.HRJ477	R---M-L------G
B_BR3.RJ483	-G--M-L-------
B_NL10.BR8914	----L-L-------
B_BR5.51	----M-V---L---
B_TH4.N76	----M-Q-------
B_BR1.7946	----L-L-----T-
B_NL1.A9	------L-----T-
B_VE.VE4	R---M-F---L-TN
B_BR3.RJ12	------W---I---
B_LT.LIT22A	R-----W---V---
B_BR3.RJI02	----V-V---I---
B_BR5.506	---QL-M---L---
B_BR6.P6	-G--M-M---L-V-
B_AR1.06	----MAG---I---
B_AR1.22	----MAG---L---
B_AR1.56	----LAG-K-L---

18

A_CF1.4058	R-----S-----II
B_GB.V12	R-----S-----TI
A_CF2.GAN	-RW---S-Q----I
B_DE.D3	-R-R--AR----TK
F_CM.CA16	RR-R--L--V----

19

G_GA.VI525	---KF-T--VL---
G_GA1.VI526	--KKF-T--VL---
B_BR3.RJ19	---QY-T-GGAFYA

20

B_TH7.181	----L---Q-W-T-
B_TH7.99	----L---Q-W-T-
B_TH7.177	----L---Q-W-T-
B_TH7.160	----L---Q-W-T-
B_TH7.151	----L---Q-W-T-
B_TH7.133	----L---Q-W-T-
B_TH7.132	----L---Q-W-T-
B_TH7.130	----L---Q-W-T-
B_TH7.128	----L---Q-W-T-
B_TH7.123	----L---Q-W-T-
B_TH7.117	----L---Q-W-T-
B_TH1.8669	----L---Q-W-T-
B_TH1.8645	----L---Q-W-T-
B_MY3.M1739	----L---Q-W-T-
B_MY3.1763	----L---Q-W-T-
B_MY3.1748	----L---Q-W-T-
B_MY2.9315174	----L---Q-W-T-
B_MY2.9315172	----L---Q-W-T-
B_MY2.9315168	----L---Q-W-T-
B_MY2.9315157	----L---Q-W-T-
B_MY2.9214096	----L---Q-W-T-
B_MY2.9214093	----L---Q-W-T-
B_MY2.9214089	----L---Q-W-T-
B_CN1.195	----L---Q-W-T-
B_CN1.193	----L---Q-W-T-
B_CN1.174	----L---Q-W-T-
B_CN1.167	----L---Q-W-T-
B_CN1.159	----L---Q-W-T-
B_CN1.157	----L---Q-W-T-
B_CN1.141	----L---Q-W-T-
B_CN1.074	----L---Q-W-T-
B_CN1.069	----L---Q-W-T-
B_CN1.068	----L---Q-W-T-
B_CN1.064	----L---Q-W-T-
B_CN1.052	----L---Q-W-T-
B_TH7.163	--------Q-W-T-
B_CN1.029	R---L---Q-W-T-
B_TH7.167	----L---Q-WFT-
B_TH2.CM237	----L---K-W-T-
B_TH6.0258	----L---K-W-T-
B_CN1.142	----L---Q-WHT-
B_TH7.162	----L-----W-T-
B_US18.LBI	----L-----W-T-
B_TH7.110	----L-----W-T-
B_TH3.W2TH014	----L-----W-T-
B_TH1.7787	----L-----W-T-
B_CN1.202	----L-----W-T-
B_TH7.168	--V-L-----W-T-
B_TH7.120	----L-----WWT-
B_CN1.161	---SL-----W-T-
B_TH7.131	---SL-----W-T-
B_CN1.171	---SL---Q-W-T-
B_CN1.049	---SL---K-WFT-
B_CN1.051	---SL---K-WFT-
B_CN1.162	---SL-----WHT-
B_TH1.8647	---PL-----W---
B_TH7.173	---PL-----W---
B_CN1.163	---PL-----W---
B_GB4.M73768520	---P------W---
B_US10.MA145con	---PM-----WF--
B_TH7.141	---QL-----W---
B_TH7.157	---QL-----W---

B_TH7.150	---Q------W---
B_NO1.23	R---------W---
B_TH7.165	R---L-----W--P
B_BR3.HRJ625	----------WF-R
B_CN.1798	----L-Q---W-T-
B_TH7.102	----L-Q---W-T-
B_TH7.183	----L-Q---W---
B_TH1.8643	----L-Q-Q-W---
B_TH7.112	---QL-Q---W-T-
B_AR1.21	---NM-A---W---
B_TH7.96	---NL-L---W---
B_TH7.90	---PL---Q-W-T-
B_TH7.94	---PL---Q-W-T-
B_TH7.88	---PL---Q-W-T-
B_TH7.87	---PL---Q-W-T-
B_TH7.86	---PL---Q-W-T-
B_TH7.85	---PL---Q-W-T-
B_TH7.82	---PL---Q-W-T-
B_TH7.153	---PL---Q-W-T-
B_TH7.137	---PL---Q-W-T-
B_TH7.122	---PL---Q-W-T-
B_TH7.113	---PL---Q-W-T-
B_TH7.104	---PL---Q-W-T-
B_TH5.2619	---PL---Q-W-T-
B_TH3.W2TH026	---PL---Q-W-T-
B_TH1.8653	---PL---Q-W-T-
B_TH1.8649	---PL---Q-W-T-
B_TH.T8174	---PL---Q-W-T-
B_TH.93TH067	---PL---Q-W-T-
B_MY3.1782	---PL---Q-W-T-
B_MY3.1755	---PL---Q-W-T-
B_MY2.9315171	---PL---Q-W-T-
B_MY2.9315160	---PL---Q-W-T-
B_CN1.170	---PL---Q-W-T-
B_CN1.147	---PL---Q-W-T-
B_JP.JNIH1M	---PL---Q-W-S-
B_TH7.182	---PL---Q-WFT-
B_TH7.92	---PL---Q-WFT-
B_TH7.166	---PL---Q-WFT-
B_TH7.116	---PL---Q-WFT-
B_TH7.105	---PL---Q-WFT-
B_TH1.8655	---PL---QVWFT-
B_MY1.MRN02	---PL---QTW-T-
B_TH7.149	---PL---KTW-T-
B_TH4.N763	-G-PL---Q-W-T-
B_TH7.124	-G-P----Q-W-T-
B_TH7.118	-G-PL-----W-T-
B_TH7.169	---PL-----W-T-
B_TH7.93	---PL-----W-T-
B_TH7.152	---PL-----W-T-
B_TH7.139	---PL-----W-T-
B_CN1.200	---PL-----W-T-
B_CN1.150	---PL-----W-T-
B_TH7.158	--VPL-----W-T-
B_MY2.9315158	---PL---K-W-T-
B_TH7.174	---PL---K-W-T-
B_MY2.9214087	---PL---K-W-T-
B_CN1.206	---PL---K-W-T-
B_CN1.199	---PL---K-W-T-
B_CN1.173	---PL---K-W-T-
B_NL.114C	---P------W-T-
B_US6.I06	---P------W-T-
B_NL10.UM893272	R--P------W-T-
B_CN1.144	---PL-S---W-T-
B_TH7.176	---PL-A---W-T-
B_TH7.109	---PL-L---W-T-
B_CN1.151	---PL-L-K-W---
B_CN1.169	-RVPL----VW-T-
B_CN1.192	-RVPL----VW-T-
B_PY.12837	-RVSL----VW-T-
B_US.CDC42	-RVTL----VW-T-
B_TH7.101	-RVTL---KVW-T-
B_CY1.HO40	-R-TM----VW-T-
B_NL1.A13	-RVTM----VW-T-
B_US11.349	--VTM----VW-T-
B_IT1.196	--VTV----VW-T-
B_TH7.108A	-GTQ-----TW-T-
B_TH7.103	-G--L---Q-W-T-
B_TH7.155	-G--L---Q-W-T-
B_TH4.N762	-G--L---Q-W-T-
B_TH1.865	-G--L---Q-W-T-
B_TH7.154	-G------Q-W-T-
B_GB5.4661	-G-------TW-T-
B_TH7.159	-G--L-----W-T-
B_TH7.129	-G--L-L---W-T-
B_GB4.AC03	-GV-L---S-W---

21

D_UG5.UG269	-TR-RAQ---WWT-

22

O_CONSENSUS_96	QE-K---MAWYSMG
O_FR.CFBCF07	QE-K---MAWYSMG
O_FR.CFBCF02	QE-R---MAWYSMG
O_FR.DUR	QE-K---MAWYSMQ
O_CM.ANT70	QEMR---MAWYSMG
O_GA1.VI686	QEMK---MAWYSMG
O_ES1.1158	QEMK---MAWYSMA
O_FR.CFBCF01	QEMK---LSWYSMG
O_FR.CFBCF08	QE--S--MAWYSLG
O_FR.VAU	QK-MA--MAWYSMA
O_CM.MVP5180	QD-YT--M-WRSM-
O_FR.CFBCF06	QR-AT--L-WVSMA

23

D_UG1.W2UG024	QRTP--L-Q-L-T-
D_UG2.3	QRTP--L-Q-L-T-

COUNTRY CODES

It is becoming increasingly useful to name viral isolates and samples with a country code. The following code was captured from Internet files:

gopher://kupe.itu.ch/11/.1/itudoc/public/gophermenus/.1/.un/.edicore/.wp4/.sept95/.rdocs95/.28024
gopher://kupe.itu.ch/00/.1/ITU-Databases/.1/CtryCodes/.full-list.txt

for ISO two-letter codes and for ITU three-letter codes, respectively.
This is a list based on the International Organization for Standardization (ISO) 3166:1993 standard, updated from a list prepared by Mark Horton. Note that the original standard has this same information sorted into about 6 different orders, both in English and French, therefore this is an abbreviated version not to be taken as the entire standard. While it has been checked against the standard, it may possibly contain errors; the standard and registration newsletters should be verified for any critical application. This copy has been updated and is believed to be current through September 1996.

Table of Country Codes from ISO 3166

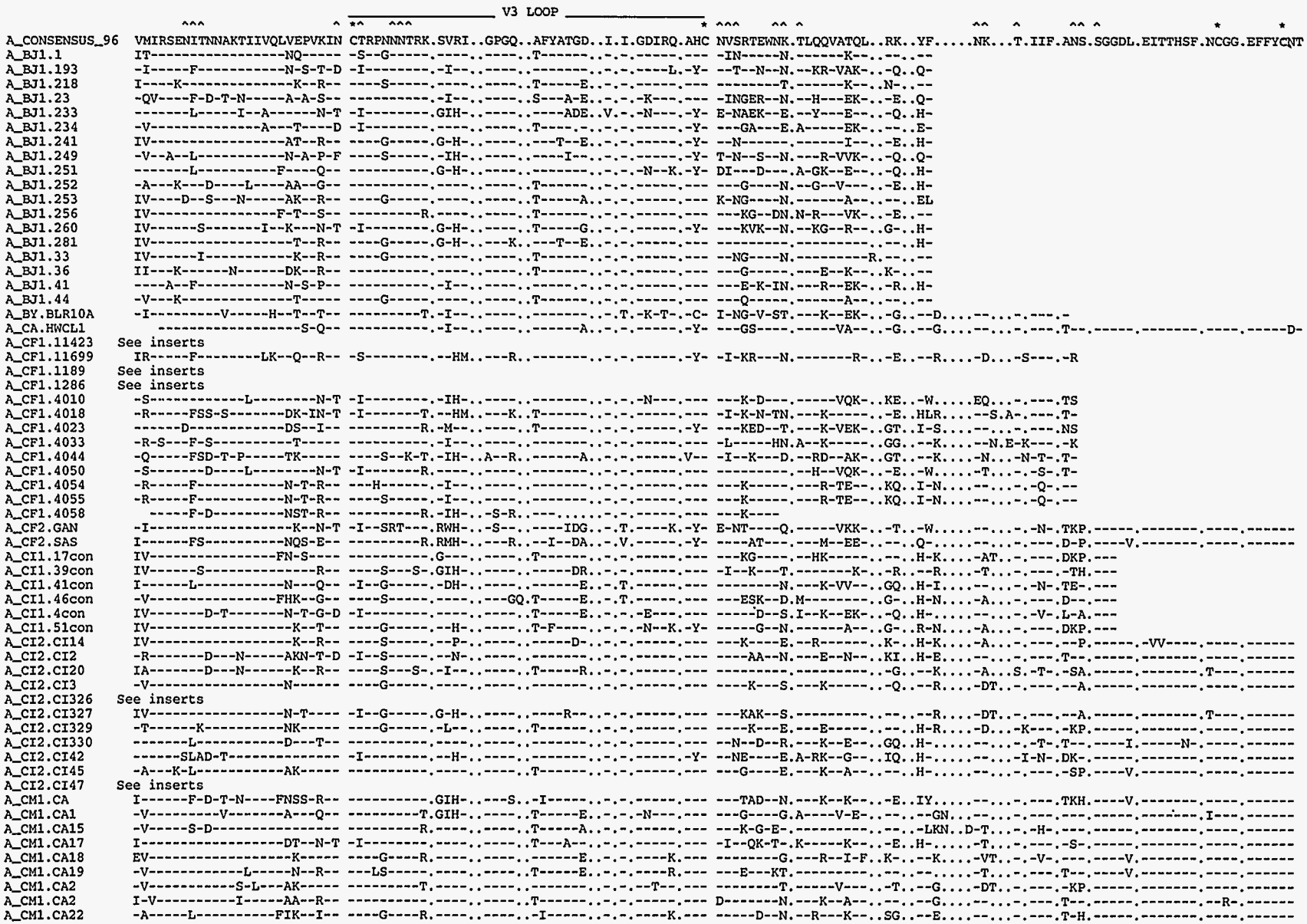
Country	A 2	A 3	Number
AFGHANISTAN	AF	AFG	004
ALBANIA	AL	ALB	008
ALGERIA	DZ	DZA	012
AMERICAN SAMOA	AS	ASM	016
ANDORRA	AD	AND	020
ANGOLA	AO	AGO	024
ANGUILLA	AI	AIA	660
ANTARCTICA	AQ	ATA	010
ANTIGUA AND BARBUDA	AG	ATG	028
ARGENTINA	AR	ARG	032
ARUBA	AW	ABW	533
AUSTRALIA	AU	AUS	036
AUSTRIA	AT	AUT	040
AZERBAIJAN	AZ	AZE	031
BAHAMAS	BS	BHS	044
BAHRAIN	BH	BHR	048
BANGLADESH	BD	BGD	050
BARBADOS	BB	BRB	052
BELGIUM	BE	BEL	056
BELIZE	BZ	BLZ	084
BENIN	BJ	BEN	204
BERMUDA	BM	BMU	060
BHUTAN	BT	BTN	064
BOLIVIA	BO	BOL	068
BOTSWANA	BW	BWA	072
BOUVET ISLAND	BV	BVT	074
BOSNIA AND HERZEGOVINA	BA	BIH	070
BRAZIL	BR	BRA	076
BRITISH INDIAN OCEAN TERRITORY	IO	IOT	086
BRUNEI DARUSSALAM	BN	BRN	096
BULGARIA	BG	BGR	100
BURKINA FASO	BF	BFA	854
BURUNDI	BI	BDI	108
BYELORUSSIAN SSR	BY	BYS	112
CAMBODIA	KH	KHM	116

CAMEROON	CM	CMR	120
CANADA	CA	CAN	124
CAPE VERDE	CV	CPV	132
CAYMAN ISLANDS	KY	CYM	136
CENTRAL AFRICAN REPUBLIC	CF	CAF	140
CHAD	TD	TCD	148
CHILE	CL	CHL	152
CHINA	CN	CHN	156
CHRISTMAS ISLAND	CX	CXR	162
COCOS (KEELING) ISLANDS	CC	CCK	166
COLOMBIA	CO	COL	170
COMOROS	KM	COM	174
CONGO	CG	COG	178
COOK ISLANDS	CK	COK	184
COSTA RICA	CR	CRI	188
COTE D'IVOIRE	CI	CIV	384
CROATIA	HR	HRV	191
CUBA	CU	CUB	192
CYPRUS	CY	CYP	196
CZECH REPUBLIC	CZ	CZE	203
DENMARK	DK	DNK	208
DJIBOUTI	DJ	DJI	262
DOMINICA	DM	DMA	212
DOMINICAN REPUBLIC	DO	DOM	214
EAST TIMOR	TP	TMP	626
ECUADOR	EC	ECU	218
EGYPT	EG	EGY	818
EL SALVADOR	SV	SLV	222
EQUATORIAL GUINEA	GQ	GNQ	226
ERITREA	ER	ERI	232
ESTONIA	EE	EST	233
ETHIOPIA	ET	ETH	230
FALKLAND ISLANDS (MALVINAS)	FK	FLK	238
FAROE ISLANDS	FO	FRO	234
FIJI	FJ	FJI	242
FINLAND	FI	FIN	246
FRANCE	FR	FRA	250
FRANCE, METROPOLITAN	FX		249
FRENCH GUIANA	GF	GUF	254
FRENCH POLYNESIA	PF	PYF	258
FRENCH SOUTHERN TERRITORIES	TF	ATF	260
GABON	GA	GAB	266
GAMBIA	GM	GMB	270
GEORGIA	GE	GEO	268
GERMANY	DE	DEU	276
GHANA	GH	GHA	288
GIBRALTAR	GI	GIB	292
GREECE	GR	GRC	300
GREENLAND	GL	GRL	304
GRENADA	GD	GRD	308
GUADELOUPE	GP	GLP	312
GUAM	GU	GUM	316
GUATEMALA	GT	GTM	320
GUINEA	GN	GIN	324

GUINEA-BISSAU	GW	GNB	624
GUYANA	GY	GUY	328
HAITI	HT	HTI	332
HEARD AND MCDONALD ISLANDS	HM	HMD	334
HONDURAS	HN	HND	340
HONG KONG	HK	HKG	344
HUNGARY	HU	HUN	348
ICELAND	IS	ISL	352
INDIA	IN	IND	356
INDONESIA	ID	IDN	360
IRAN (ISLAMIC REPUBLIC OF)	IR	IRN	364
IRAQ	IQ	IRQ	368
IRELAND	IE	IRL	372
ISRAEL	IL	ISR	376
ITALY	IT	ITA	380
JAMAICA	JM	JAM	388
JAPAN	JP	JPN	392
JORDAN	JO	JOR	400
KAZAKHSTAN	KZ	KAZ	398
KENYA	KE	KEN	404
KIRIBATI	KI	KIR	296
KOREA, DEMOCRATIC PEOPLE'S REPUBLIC OF	KP	PRK	408
KOREA, REPUBLIC OF	KR	KOR	410
KUWAIT	KW	KWT	414
KYRGYZSTAN	KG	KGZ	417
LAO PEOPLE'S DEMOCRATIC REPUBLIC	LA	LAO	418
LATVIA	LV	LVA	428
LEBANON	LB	LBN	422
LESOTHO	LS	LSO	426
LIBERIA	LR	LBR	430
LIBYAN ARAB JAMAHIRIYA	LY	LBY	434
LIECHTENSTEIN	LI	LIE	438
LITHUANIA	LT	LTU	440
LUXEMBOURG	LU	LUX	442
MACAU	MO	MAC	446
MADAGASCAR	MG	MDG	450
MALAWI	MW	MWI	454
MALAYSIA	MY	MYS	458
MALDIVES	MV	MDV	462
MALI	ML	MLI	466
MALTA	MT	MLT	470
MARSHALL ISLANDS	MH	MHL	584
MARTINIQUE	MQ	MTQ	474
MAURITANIA	MR	MRT	478
MAURITIUS	MU	MUS	480
MAYOTTE	YT	MYT	175
MEXICO	MX	MEX	484
MICRONESIA	FM	FSM	583
MOLDOVA	MD	MDA	498
MONACO	MC	MCO	492
MONGOLIA	MN	MNG	496
MONTSERRAT	MS	MSR	500
MOROCCO	MA	MAR	504
MOZAMBIQUE	MZ	MOZ	508

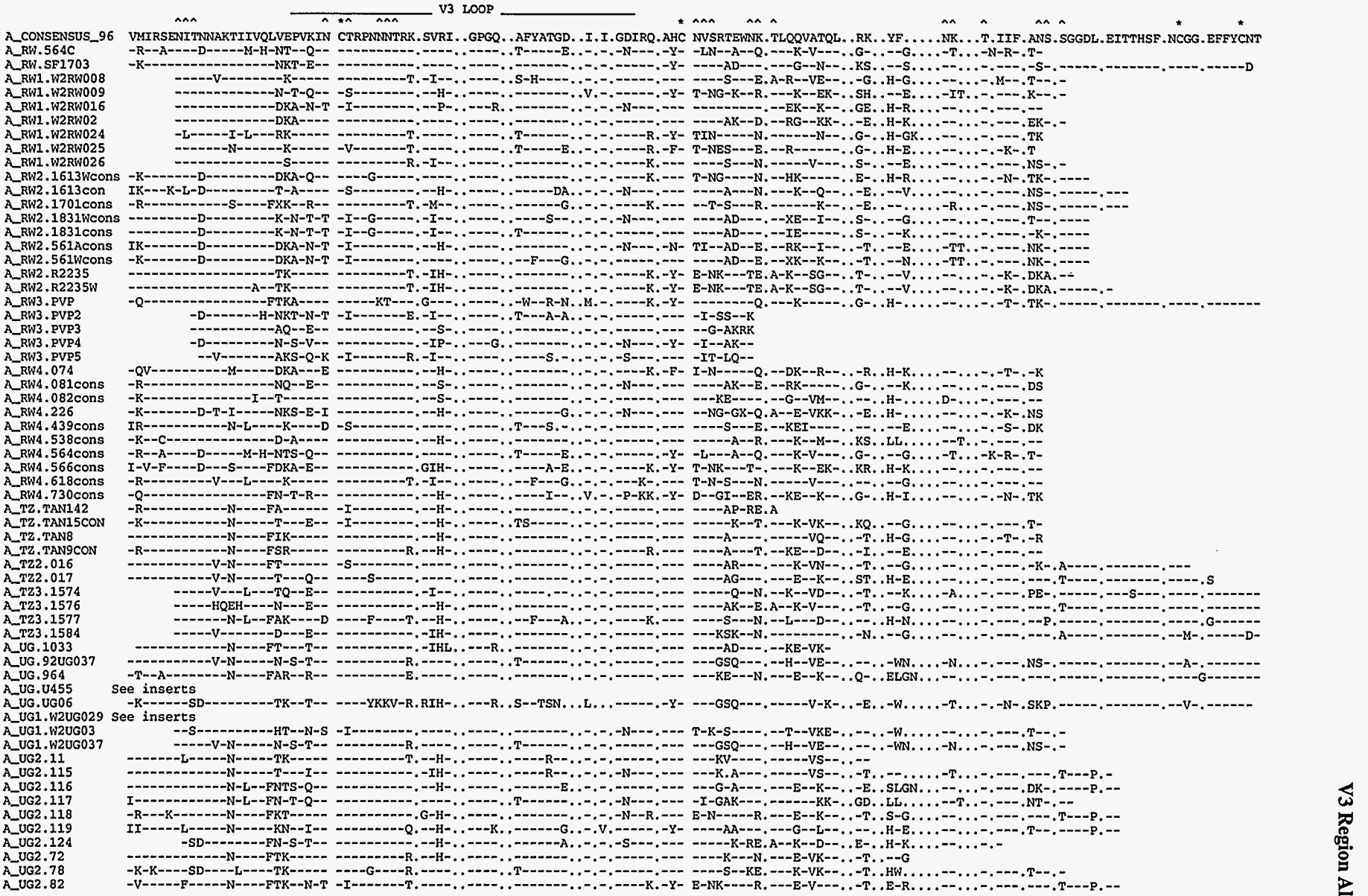
MYANMAR	MM	MMR	104
NAMIBIA	NA	NAM	516
NAURU	NR	NRU	520
NEPAL	NP	NPL	524
NETHERLANDS	NL	NLD	528
NETHERLANDS ANTILLES	AN	ANT	532
NEUTRAL ZONE	NT	NTZ	536
NEW CALEDONIA	NC	NCL	540
NEW ZEALAND	NZ	NZL	554
NICARAGUA	NI	NIC	558
NIGER	NE	NER	562
NIGERIA	NG	NGA	566
NIUE	NU	NIU	570
NORFOLK ISLAND	NF	NFK	574
NORTHERN MARIANA ISLANDS	MP	MNP	580
NORWAY	NO	NOR	578
OMAN	OM	OMN	512
PAKISTAN	PK	PAK	586
PALAU	PW	PLW	585
PANAMA	PA	PAN	590
PAPUA NEW GUINEA	PG	PNG	598
PARAGUAY	PY	PRY	600
PERU	PE	PER	604
PHILIPPINES	PH	PHL	608
PITCAIRN	PN	PCN	612
POLAND	PL	POL	616
PORTUGAL	PT	PRT	620
PUERTO RICO	PR	PRI	630
QATAR	QA	QAT	634
REUNION	RE	REU	638
ROMANIA	RO	ROM	642
RUSSIAN FEDERATION	U	RUS	643
RWANDA	RW	RWA	646
ST. HELENA	SH	SHN	654
SAINT KITTS AND NEVIS	KN	KNA	659
SAINT LUCIA	LC	LCA	662
ST. PIERRE AND MIQUELON	PM	SPM	666
SAINT VINCENT AND THE GRENADINES	VC	VCT	670
SAMOA	WS	WSM	882
SAN MARINO	SM	SMR	674
SAO TOME AND PRINCIPE	ST	STP	678
SAUDI ARABIA	SA	SAU	682
SENEGAL	SN	SEN	686
SEYCHELLES	SC	SYC	690
SIERRA LEONE	SL	SLE	694
SINGAPORE	SG	SGP	702
SLOVAKIA	SK	SVK	703
SLOVENIA	SI	SVN	705
SOLOMON ISLANDS	SB	SLB	090
SOMALIA	SO	SOM	706
SOUTH AFRICA	ZA	ZAF	710
SOUTH GEORGIA AND THE SOUTH SANDWICH ILANDS	GS		239
SPAIN	ES	ESP	724
SRI LANKA	LK	LKA	144

SUDAN	SD	SDN	736
SURINAME	SR	SUR	740
SVALBARD AND JAN MAYEN ISLANDS	SJ	SJM	744
SWAZILAND	SZ	SWZ	748
SWEDEN	SE	SWE	752
SWITZERLAND	CH	CHE	756
SYRIAN ARAB REPUBLIC	SY	SYR	760
TAIWAN, PROVINCE OF CHINA	TW	TWN	158
TAJIKISTAN	TJ	TJK	762
TANZANIA, UNITED REPUBLIC OF	TZ	TZA	834
THAILAND	TH	THA	764
TOGO	TG	TGO	768
TOKELAU	TK	TKL	772
TONGA	TO	TON	776
TRINIDAD AND TOBAGO	TT	TTO	780
TUNISIA	TN	TUN	788
TURKEY	TR	TUR	792
TURKMENISTAN	TM	TKM	795
TURKS AND CAICOS ISLANDS	TC	TCA	796
TUVALU	TV	TUV	798
UGANDA	UG	UGA	800
UKRAINE	UA	UKR	804
UNITED ARAB EMIRATES	AE	ARE	784
UNITED KINGDOM	GB	GBR	826
UNITED STATES	US	USA	840
UNITED STATES MINOR OUTLYING ISLANDS	UM	UMI	581
URUGUAY	UY	URY	858
UZBEKISTAN	UZ	UZB	860
VANUATU	VU	VUT	548
VATICAN CITY STATE (HOLY SEE)	VA	VAT	336
VENEZUELA	VE	VEN	862
VIET NAM	VN	VNM	704
VIRGIN ISLANDS (BRITISH)	VG	VGB	092
VIRGIN ISLANDS (U.S.)	VI	VIR	850
WALLIS AND FUTUNA ISLANDS	WF	WLF	876
WESTERN SAHARA	EH	ESH	732
YEMEN, REPUBLIC O	YE	YEM	887
YUGOSLAVIA	YU	YUG	890
ZAIRE	ZR	ZAR	180
ZAMBIA	ZM	ZMB	894
ZIMBABWE	ZW	ZWE	716


V3 LOOP
A_CONSENSUS_96 VMIRSENITNNAKTIIVQLVEPVKIN CTRPNNNTRK.SVRI..GPGQ..AFYATGD..I.I.GDIRQ.AHC NVSRTEWNK.TLQQVATQL..RK..YF.....NK...T.IIF.ANS.SGGDL.EITTHSF.NCGG.EFFYCNT
A_BJ1.1 IT----------NQ----- -S--G-----.----..----..T------..-.-.-----.--- -IN-----N.------K--..--..--
A_BJ1.193 -I-----F----------N-S-T-D -I--------.-I--..----..-------..-.-.----L.-Y- --T--N--N.--KR-VAK-..-Q..Q-
A_BJ1.218 I----K----------K--R-- ----S-----.----..----..T-----E..-.-.-----.--- --------T.------K--..N-..--
A_BJ1.23 -QV----F-D-T-N-----A-A-S-- ----------.-I--..----..S---A-E..-.-.-K---.--- -INGER--N.--H---EK-..-E..Q-
A_BJ1.233 -------L-----I--A------N-T -I--------.GIH-..----..----ADE..V.-.-N---.-Y- E-NAEK--E.--Y---E--..-Q..H-
A_BJ1.234 -V----------A---T----D -I--------.----..----..T----.-..-.-.-----.-Y- ---GA---E.A-----EK-..--..E-
A_BJ1.241 IV----------AT--R-- ----G-----.G-H-..----..---T--E..-.-.-----.-Y- --N------.------I--..-E..H-
A_BJ1.249 -V--A--L----------N-A-P-F ----S-----.-IH-..----..----I--..-.-.-----.-Y- T-N--S--N.---R-VVK-..-Q..Q-
A_BJ1.251 -------L----------F----Q-- ----------.G-H-..----..-------..-.-.-N--K.-Y- DI---D---.A-GK--E--..-Q..H-
A_BJ1.252 -A---K---D----L----AA--G-- ----------.----..----..T------..-.-.-----.--- ---G----N.--G--V----..-E..H-
A_BJ1.253 IV----D--S---N-----AK--R-- ----G-----.----..----..T-----A..-.-.-----.--- K-NG----N.------A--..--..EL
A_BJ1.256 IV----------F-T--S-- ---------R.----..----..T------..-.-.-----.--- ---KG--DN.N-R----VK-..-E..--
A_BJ1.260 IV------S-------I--K---N-T -I--------.G-H-..----..T-----G..-.-.-----.-Y- ---KVK--N.--KG--R---..G-..H-
A_BJ1.281 IV----------T--R-- ----G-----.G-H-..---K..---T--E..-.-.-----.--- ---------.---------..--..H-
A_BJ1.33 IV------I----------K--R-- ----G-----.----..----..T------..-.-.-----.--- --NG----N.---------R.--..--
A_BJ1.36 II---K------N------DK--R-- ----------.----..----..T------..-.-.-----.--- ---G-----.---E--K---..K-..--
A_BJ1.41 ----A--F----------N-S-P-- ----------.-I--..----..-----.-..-.-.-----.--- ---E-K-IN.---R--EK-..-R..H-
A_BJ1.44 -V---K----------T----- ----G-----.----..----..T------..-.-.-----.--- ---Q-----.------A--..--..--
A_BY.BLR10A -I---------V-----H--T--T-- ---------T.-I--..----..-------..-.T.-K-T-.-C- I-NG-V-ST.---K--EK-..-G..--D....--...-.---.-
A_CA.HWCL1 ----------S-Q-- ----------.-I--..----..------A..-.-.-----.-Y- ---GS----.-----VA---..-G..--G....--...-.---.T--.-----.-------.----.-----D-
A_CF1.11423 See inserts
A_CF1.11699 IR-----F--------LK--Q--R-- -S--------.--HM..---R..-------..-.-.-----.-Y- -I-KR---N.-------R-..-E..--R....-D...-S---.-R
A_CF1.1189 See inserts
A_CF1.1286 See inserts
A_CF1.4010 -S----------L--------N-T -I--------.-IH-..----..-------..-.-.-N---.--- ---K-D---.-----VQK-..KE..-W.....EQ...-.---.TS
A_CF1.4018 -R-----FSS-S-------DK-IN-T -I-------T.--HM..---K..T------..-.-.-----.--- -I-K-N-TN.---K-----..-E..HLR....--S.A-.---.T-
A_CF1.4023 ------D----------DS--I-- ---------R.-M--..----..T------..-.-.-----.-Y- ---KED--T.---K-VEK-..GT..I-S....--...-.---.NS
A_CF1.4033 -R-S---F-S----------T----- ----------.-I--..----..-------..-.-.-----.--- -L-----HN.A--K------..GG..--K.....-N.E-K---.-K
A_CF1.4044 -Q-----FSD-T-P-----TK----- ----S--K-T.-IH-..A--R..------A..-.-.-----.V-- -I--K---D.--RD--AK-..GT..--K....-N...-N-T-.T-
A_CF1.4050 -S-------D----L--------N-T -I--------R.----..----..-------..-.-.-----.--- ---------.--H--VQK-..-E..-W.....-T...-.-S-.T-
A_CF1.4054 -R-----F----------N-T-R-- ---H-------.-I--..----..-------..-.-.-----.--- ---K-----.---R-TE---.KQ..I-N....--...-.-Q-.--
A_CF1.4055 -R-----F----------N-T-R-- ----S-----.-I--..----..-------..-.-.-----.--- ---K-----.---R-TE---.KQ..I-N....--...-.-Q-.--
A_CF1.4058 -----F-D----------NST-R-- ---------R.-IH-..-S-R..---.....-.-.-----.--- ---K----
A_CF2.GAN -I----------K--N-T -I--SRT---.RWH-..-S--..----IDG..-.T.----K.-Y- E-NT----Q.-----VKK-..-T..-W.....--...-.-N-.TKP.-----.-------.----.------
A_CF2.SAS I------FS----------NQS-E-- ---------R.RMH-..---R.--I--DA..-.V.------.-Y- ----AT---.---M--EE-..--..Q-.....--...-.---.D-P.----V.-------.----.------
A_CI1.17con IV----------FN-S---- ----------.G---..----..T------..-.-.-----.--- ---KG-----.--HK-----..--..H-K....-AT..-.---.DKP.---
A_CI1.39con IV------S----------R-- ----S---S-.GIH-..----..-----DR..-.-.-----.--- -I--K---T.-------K-..-R..--R....-T...-.---.-TH.---
A_CI1.41con I------L----------N---Q-- -I--G-----.-DH-..----..------E..-.T.-----.--- --------N.---K-VV---.GQ..H-I....--...-.-N-.TE-.---
A_CI1.46con -V----------FHK--G-- ----S-----.----..---GQ.T-----E..-.T.-----.--- ---ESK--D.M--------.G-..H-N....-A...-.---.D--.---
A_CI1.4con IV-------D-T-------N-T-G-D -I--------.----..----..T-----E..-.-.-E----.--- -----D--S.I--K--EK-..-Q..H-.....--...-.-V-.L-A.---
A_CI1.51con IV----------K--T-- ----G-----.--H-..----..T-F----..-.-.-N--K.-Y- -----G--N.------A---.G-..R-N....-A...-.---.DKP.---
A_CI2.CI14 IV----------K--R-- ----S-----.--P-..----..-----D-..-.-.-----.--- ---K----E.--R-------.K-..H-K....-A...-.---.--P.-----.-VV----.----.------
A_CI2.CI2 -R-------D---N-----AKN-T-D -I--S-----.--N-..----..-------..-.-.-----.--- ----AA--N.---E--N---..KI..H-E.....--...-.---.T--.-----.-------.----.------
A_CI2.CI20 IA-------D---N------K--R-- ----S---S-.-I--..----..T-----R..-.-.-----.--- ---------.---------..-G..--K....-A...S.-T-.-SA.-----.-------.T---.------
A_CI2.CI3 -V----------N------ ----G-----.----..----..-------..-.-.-----.--- ----K---S.---K------..-Q..--R....-DT..-.---.--A.-----.-------.----.------
A_CI2.CI326 See inserts
A_CI2.CI327 IV----------N-T---- -I--G-----.G-H-..----..----R--..-.-.-----.--- ---KAK--S.---------..--..--R....-DT..-.---.--A.-----.-------.T---.------
A_CI2.CI329 -T------K----------NK----- ----G-----.-L--..----..T------..-.-.-----.--- ----K---E.---E-----..--..H-R....-D...-K---.-KP.-----.-------.----.------
A_CI2.CI330 -------L----------D---T-- ----------.----..----..-------..-.-.-----.--- --N--D--R.---K--E--..GQ..H-.....--...-.-T-.T--.----I.-----N-.----.------
A_CI2.CI42 ------SLAD-T---------- -I--------.--H-..----..-------..-.-.-----.-Y- --NE----E.A-RK--G---.IQ..H-.....--...-I-N-.DK-.-----.-------.----.------
A_CI2.CI45 -A---K-L----------AK----- ----------.----..----..T------..-.-.-----.--- ---G----E.---K--A---..--..H-.....--...-.---.-SP.----V.-------.----.------
A_CI2.CI47 See inserts
A_CM1.CA I------F-D-T-N----FNSS-R-- ----------.GIH-..--S..-I-----..-.-.-----.--- ---TAD--N.---K--K---..-E..IY.....--...-.---.TKH.----V.-------.----.------
A_CM1.CA1 -V---------V-------A---Q-- ---------T.GIH-..-----..T-----E..-.-.-N---.--- ---G----G.A----V-E-..--..--GN...--...-.---.----.-----.---.I---.------
A_CM1.CA15 -V-----S-D---------------- ---------R.----..----..T-----A..-.-.-----.--- ---K-G-E-.----------..--..-LKN..D-T...-.-H-.---.-----.-------.----.------
A_CM1.CA17 I------------------DT--N-T -I--------.----..----..T---A--..-.-.-----.--- -I--QK-T-.K-----K---..-E..H-.....-T...-.---.-S-.-----.-------.----.------
A_CM1.CA18 EV----------K----- ----G----R.----..----..------E..-.-.----K.--- --------G.---R--I-F..K-...-K....-VT..-.-V-.---.----V.-------.----.------
A_CM1.CA19 -V----------L-----N--R-- ---LS-----.----..----..T-----E..-.-.----R.--- ---E---KT.----------..--..-----T...-.---.T--.----V.-------.----.------
A_CM1.CA2 -V----------S-L---AK----- ---------T.----..----..-------..-.-.--T--.--- --------T.-----V----..-T..--G....-DT..-.---.-KP.-----.-------.----.------
A_CM1.CA2 I-V----------I-----AA--R-- ----------.----..----..T------..-.-.-----.--- D-------N.---K--A---..--..--G....--...-.---.T--.-----.-------.--R-.------
A_CM1.CA22 -A-----L----------FIK--I-- ----G----R.----..----..-I-----..-.-.----K.--- -----D--N.--R---K---..SG..--E.....--...-.---.T-H.-----.-------.----.------

V3 LOOP

```
                  ^^^           ^ *^   ^^^                                  * ^^^    ^^ ^                      ^^   ^      ^^ ^              *       *
A_CONSENSUS_96    VMIRSENITNNAKTIIVQLVEPVKIN CTRPNNNTRK.SVRI..GPGQ..AFYATGD..I.I.GDIRQ.AHC NVSRTEWNK.TLQQVATQL..RK..YF.....NK...T.IIF.ANS.SGGDL.EITTHSF.NCGG.EFFYCNT
A_CM1.CA6         IR-----F-D-----L---NST-R-- -S-------R.-IHF..-----..TL----N..-.-.-----.--- ----K--EN.M----T-H-..KE..K-.....--...-K-T-.KS-.-----.-------.--R-.-------
A_CM1.CA7         -RV----F-D-----L---NST-R-- ---------R.-I--..-S--..TS-----..-.-.-N--E.--- -------ED.M------H-..GG..K-.....-T...-R---.KSP.-----.-------.--R-.-------
A_CY.HOcon        I---------D---N----FT---P-- ----------.-I-F..-----..---TN.-..-.-.-N---.--- -I-KKL--A.--RK--E--..G-..Q-P....--...-.---.T--
A_DJ.DJ258        -V----------N------T--R-- -S--G-----.-----..-----..T-------..-.-.-----.--- ----SK--E.---K-----..---..H-.....-T...-.---.---.----I.-------.--R-.-------
A_DJ.DJ263        -V----------QN------A--R-- ----------.-----..-----..T-------..-.-.-----.--- ----S---R.---------..---..H-.....--...-.---.---.----I.-------.-----.-------
A_DJ.DJ264        -V----------N------T--G-- ----------.-----..-----..T-------..-.-.----K.--- ----SK--N.------I---..---..HS.....-T...-.---.--P.----I.-------.----.-------
A_GA.VI19         I------F-D-G------------- ----S-----.GIH-..---R..------Q..-.T.-----.--- -I--------.--H---I--..-E..H-IT..K-T...-.---.NS-.-----.-------.----.-----S.
A_GA1.G135        -V-----L-----------GQ--I-- ---------R.GIH-..-----..--------..-.-.----K.-Y- T-N--A--E.---RDKVK-..GE..--N....--...-.-N-.NS-.-----.-------.----.--------
A_GA1.G41         -------L-----------GQ--I-- ----------.-IH-..-----..--------..-.-.-----.--- I-NG----E.--RK--KK-..-T..--E....--...-.---.T-P.-----.------V.----.--------
A_GA1.LBV23       -I----D----------FEQ--R-- ---------R.EIH-..-----.-----.-N..-.-.-----.-Y- T---AT----.--YK---K-..KT..--K....--...-.---.NE-.-----.-------.--R-.-------
A_GA1.LBV2310     -V-----------------T--R-- ----------.-----..-----..T-------..-.-.-----.--- ----A---T.------V---..--..--K....-TT..-.-T-.--P.----V.---M----.--R-.--------
A_GA1.VI1076      IA----------------D---R-- ----S-----.-----..-----..T-------..-.-.-----.--- ---GI---N.--HK-VA--..G-..--K....--...-.-N-.TKP.-----.-------.I---.--------
A_GA1.VI685       -------L----------NK--E-- ----------.--H-..-----..--------..-.-.-----.--- ----AK--E.--RK-TEK-..KT..--.....--...-.---.DTP.-----.-------.V-----.--------
A_GH.D687         IV----------R-------K--R-- ----G---N-.N-H-..-----..----R-R..-.-.----K.--- ----A---R.-------EH-..--..--K....-N...-.---.-HH.-----.-------.--R-.-------
A_IN1.CMCH14      -R------SD---------S---T-- ----------.-IH-..-----..--------..-.-.-----.--- ---G-K---.---D-----..-
A_IN1.CMCH9       -K----------------AK----S ----------.-IK-..-----..T-------..-.-.E----.-Y- E-N-S-X--.-----GE--..G
A_KE.K89          ------------N----FA------ ---------M.-I--..-----..--------..-.-.-N----.--- -----A---T.---K-V-K-..-E..--GN...--...-.-K-.---.-----.-------.----.------
A_KE.KEN112       -K----------------F-N--T-- ----------.-IH-..-----..--------..-.-.-N---.--- -----A---.---E-----..--..H-
A_KE.KEN29        -I----------------------Q-- ---------T.GIH-..-----.----R-A..-.T.----K.-Y- ----SA----.--GK--M--..--..--
A_KE.KEN3         II----------------D---I-- ----X-----.-IH-..-----..--------..-.-.----K.--- ----S---N.---K-----..--..--
A_KE.KEN88        -T---------I-------AT--T-- ---------T.-I--..-----..T-------..-.-.-N---.--- ----SA--A.---K--K---..GQ..--
A_KE.KEN965       -K-----F---V---L----K----- ----G----T.-I--..-----..---GR-N..-.-.----K.--- ---XS---N.---E--E--..KE..H-
A_KE.KEN967       -R------SD-----L---TT--A-- ---------R.-I--..-----..--------..-.T.-----.--- ----SQ--E.--RE-VI--..--..HW
A_KE.KEN968       IK---------V---------N--T-- -A-------R.-IH-..-----..----S.-..-.-.-----.--- ----S-----.V-K----H-..--..H-
A_KE.KEN970       IR----------N----FDN----- ----------.--H-..-----..--------..-.-.-----.--- -----A--E.---K--K---..KT..--
A_KE.KEN973       LI-----L-----------KT----- ----------.-I--..-----..---T----..-.-.-----.--- -I--SD--T.---K--Q---..--..X-
A_KE.KEN977       -R------SD--------FKT--T-- -I-------T.-I--..-----..--------..-.-.-----.--- ----S---E.-----VK--..--..HW
A_KE.KEN978       -S----------N----FT------- ----------.--P-..-----..-----DN..-.-.-----.--- ----V---N.----R--K---..-T..--
A_KE.KEN979       -K---------V-I-----AK----- ---------T.-IC-..-----..T---.-G..-.-.-----.--- ---E----D.N-K--T----..K-..H-
A_KE.KEN98        -K----------N----FNN----- ----------.--H-..-----..--------..-.-.-N---.--- -----A--N.---K-VS---..-T..--
A_KE.KEN98        EI---------V-------------- ----S----T.-I--..-----..--------..-.-.-----.--- ----S-----.---K--D---..-T..--
A_KE.KEN980       -K------------------K----- ---------X.-IH-..-----..--------..-.-.-----.--- ----S-----.-----G----..-T..--
A_KE.KEN982       I----------V-N-----N-S-T-- ----------.-AN-..-----..------E..-.-.-----.--- ----A-----.A-RE--R---..-T..V-
A_KE.KEN983       -K----------N-L--FN-T----- -I------L-.GIH-..-----..S-----S..-.V.-N---.--- ---GS-----.S--K-V----..-N..--
A_KE.KEN984       -I----------N------T--T-D -I--Y-----.-IH-..-----..----S.R..-.-.-N---.--- --T-SN--N.A-G---D---..-Q..--
A_KE.KEN985       -V---------V----------I-- ----------.-I--..-----..T---.-E..-.-.-N----.--- ----SK--N.---K-----..-E..H-
A_KE1.NA111       -R----------N-----TK--R-- ----------.--H-..-----..--------..-.-.-N----.--- V-NG-----.--GE--K---..-T..--Q....--...-.---.T--
A_KE1.NA112       IK----------------FDK--I-- ----S----S.-I--..-----..------A..-.-.-----.-S- T-NGS---T.---K-GEK-..-T..--V....--...-.-T-.---
A_KE1.NA114       ------------RN----FK---A-- ----------.--H-..-----..--------..-.-.-----.--- -----K-KE.-------IR-..G-..--K....--...-.---.---
A_KE1.NA115       IK----------------------K----- -I------KT.-IS-..---K..--F-S.-..-.-.-N---.--- ---KLD--T.---E--Q---..E-..--R....--...-.-K-.---
A_KE1.NA117       -R---------V---------D----- ----S----T.-----..-----..T-----.-..-.-.-----.--- ----L-----.A----VE---..T-..--S....--...-.---.T--
A_KE1.NA118       -I-------D-V-------DK--I-- ----------.-I--..-----..---TS.N..-.-.-----.--- ----SL-DE.--RK---K-..GE..--K....--...-.-N-.NK-
A_NG.IBNG         -V----------------AN----- ----------.G-H-..-----..--------..-.-.-----.--- ---K------.--H--V----..KT..--K....-T...-.---.--P.L---V.-------.-----.------
A_NL1.AO9246      I------FS----------AQ--Q-- ---------R.--H-..-----..------A..V.-.-N---.--- -I-K-G-TE.--R---EK-..G-..VS.....-I...-A---
A_NL1.GH9023      IV-----L-----------A---G-- ----S----R.-I--..-----..------G..-.-.-----.--- --NG----T.----------..-E..H-K....DS...-.---
A_NL1.GH9152      ------DL----------T------ ----SD-R--.-----..-----..T-------..-.-.-----.--- --N--K--N.------I---..---..--K....-TT..-.---
A_NL1.GH9299      -V----------------QK--T-K --T--T-----.G---..-----..V----DE..-.-.-----.--- --------R.--H---P---..--..--.....-T...-.---
A_NL1.GH94014     -V------S----I---------R-- ---------R.-----..-----..T-------..-.-.-----.-Y- E-NE-V---.--Y-------..--..--N....-N...-.---
A_NL1.GH95012     IV----------------DK--R-- ----G----A.-IHM..-----..----P--..-.-.-----.--- ----KD--D.---R------..S-..H-E....-A...-N---
A_NL1.KE9135      -R---------V---------K--T-- -----D-I--.-I--..-----..--------..-.T.-----.--- ----S---D.------G---..-E..--K....--...-.---
A_NL1.RW8903      -R---------V-N-----D---E-T ----------.-----..-----..T-------..-.-.-----.--- ---GAK--N.---E--K---..-G..--E....-R...-.---
A_NL1.RW8935      -K----------N----FHQ--E-- ---------Q.--H-..-----..-------E..-.G.------.-Y- TI---K--E.P--E--K---..QN..--E....-T...-.---
A_NL1.TZ925825    -R------SD---------T---T-- ---------T.-I--..-----..------A..-.-.-----.--- ----SR--E.--EK-VV---..--..HW.....--...-.---
A_NL1.TZ94016     ------------N----FNKA-Q-- -S---------.--H-..-----..------G..-.-.-----.--- -----Q--N.---K-IK---..ET..--R....--...-.-V-
A_NL1.UG9283      -R----------N-L----K----- ---------R.RIP-..-----..-H-R-N..V.-.----K.--- ---SS--S-.-------AA-..G-..H-G....-T...-V-G-
A_NL1.ZR93033     -AV---------------FT---R-- -S---------.-I--..-----..T-----G..-.-.-----.--- ----MD---.---------..YS..--N....-P...-.MV
A_RU1.GAN1                  --------N--E-- -S---------.GIH-..-----..--------..-.-.-----.--- -I---R--Q.---
A_RU1.IVA6                  -------N-T----- ----G-----.-----..-----..T-----A..-.-.-----.--- --T--N--N.---
A_RU1.MLY10       -I---K--SE-TNN-----KT--H-- -I-LA-----.-IH-..-----..-I-TN.-..-.-.-----.--- -I-EGL--D.--YEI----..--..H-N....--...-.---.KP-.T---I.-------.-----.-------
A_RU1.SHL9            K---E-TNN-----KT--Q-- ----G-----.-IH-..-----..-I-..--E..-.-.-N--R.--- -I-AGL--E.--YE-----..--..H-N....--...-.-S-.KP-.T---I.--A-
```


V3 LOOP
A_CONSENSUS_96 VMIRSENITNNAKTIIVQLVEPVKIN CTRPNNNTRK.SVRI..GPGQ..AFYATGD..I.I.GDIRQ.AHC NVSRTEWNK.TLQQVATQL..RK..YF.....NK...T.IIF.ANS.SGGDL.EITTHSF.NCGG.EFFYCNT
A_RW.564C -R--A----D-----M-H-NT--Q-- ----------.----..-----..T-----E..-.-.-N---.-Y- -LN--A--Q.---K-V----.G-..--G....-T...-N-R-.T-
A_RW.SF1703 -K----------------NKT-E-- ----------.----..-----..-------..-.-.------.-Y- ----AD---.---G--N---.KS..--S....--...-.---.-S-.------.--------.----.-----D
A_RW1.W2RW008 -----V--------K----- ---------T.-I--..-----..S-H----..-.-.------.--- ----S---E.A-R--VE---..-G..H-G....--...-.M--.T--.-
A_RW1.W2RW009 ------------N-T-Q-- -S--------.--H-..----..--------.V.-.------.-Y- T-NG-K--R.---K--EK-..SH..--E....-IT..-.---.K--.-
A_RW1.W2RW016 ------------DKA-N-T -I--------.--P-..---R..--------..-.-.-N---.--- ----------.--EK--K---.GE..H-R....--...-.---.--
A_RW1.W2RW02 ------------DKA---- ----------.----..-----..--------..-.-.------.--- ----AK--D.--RG--KK-...-E..H-K....--...-.---.EK-.-
A_RW1.W2RW024 -L-----I-L---RK----- ---------T.----..-----..T-------..-.-.----R.-Y- TIN-----N.------N---.G-..H-GK...--...-.---.TK
A_RW1.W2RW025 -------N------K----- -V-------T.----..-----..T-----E..-.-.----R.-F- T-NES---E.--R-------..G-..H-E....--...-.-K-.T
A_RW1.W2RW026 ------------S----- ---------R.-I--..-----..--------..-.-.----K.--- ----S---N.-----V----..S-..--E....--...-.---.NS-.-
A_RW2.1613Wcons -K-------D---------DKA-Q-- ----G-----.----..-----..--------..-.-.----K.--- T-NG----N.--HK------..E-..H-R....--...-.-N-.TK-.-----
A_RW2.1613con IK---K-L-D---------T-A---- -S---------.--H-..-----..-----DA..-.-.-N---.--- ----A---N.---K--Q---..-E..--V....--...-.---.NS-.------.---
A_RW2.1701cons -R----------S----FXK--R-- ---------T.-M--..-----..-------G..-.-.-----K.--- --T-S---R.------K--...-E..--.....-R...-.---.NS-.-----.---
A_RW2.1831Wcons ---------D---------K-N-T-T -I--G-----.-I--..-----..----S--..-.-.-N---.--- ----AD---.--XE--I---.S-..--G....--...-.---.T--.----
A_RW2.1831cons ---------D---------K-N-T-T -I--G-----.-I--..-----..T-------..-.-.-----.--- ----AD---.--IE------..S-..--K....--...-.---.-K-.----
A_RW2.561Acons IK-------D---------DKA-N-T -I---------.--H-..-----..--------..-.-.-N---.-N- TI--AD--E.--RK--I---..-T..--E....-TT..-.---.NK-.----
A_RW2.561Wcons -K-------D---------DKA-N-T -I---------.----..-----..--F---G..-.-.-----.--- ----AD--E.--XK--K---..-T..--N....-TT..-.---.NK-.----
A_RW2.R2235 ----------TK----- ---------T.-IH-..-----..--------..-.-.----K.-Y- E-NK---TE.A-K--SG---..T-..--V....--...-.-K-.DKA.--
A_RW2.R2235W ----------A--TK----- ---------T.-IH-..-----..--------..-.-.----K.-Y- E-NK---TE.A-K--SG---..T-..--V....--...-.-K-.DKA.-----.-
A_RW3.PVP -Q----------FTKA---- -----KT---.G---..-----..-W--R-N..M.-.----K.-Y- --------Q.---K------..G-..H-.....--...-.-T-.TK-.-----.-------.----.-------
A_RW3.PVP2 -D-------H-NKT-N-T -I--------E.-I--..-----..T---A-A..-.-.------.--- -I-SS--K
A_RW3.PVP3 ----------AQ--E-- ----------.--S-..-----..--------..-.-.------.--- --G-AKRK
A_RW3.PVP4 -D---------N-S-V-- ----------.-IP-..---G..--------..-.-.-N---.-Y- -I--AK--
A_RW3.PVP5 --V-------AKS-Q-K -I--------R.-I--..-----..----S.-..-.-.-S---.--- -IT-LQ--
A_RW4.074 -QV----------M-----DKA---E ----------.--H-..-----..--------..-.-.----K.-F- I-N-----Q.--DK--R--..-R..H-K....--...-.-T-.-K
A_RW4.081cons -R----------------NQ--E-- ----------.--S-..-----..--------..-.-.-N---.--- ----AK--E.--RK------..G-..--K....--...-.----.DS
A_RW4.082cons -K-------------I--T------ ----------.--S-..-----..--------..-.-.-----.--- ---KE-----.--G--VM---..---..H-.....D-...-.---.--
A_RW4.226 -K-------D-T-I-----NKS-E-I ----------.--H-..-----..------G..-.-.-N---.--- --NG-GX-Q.A--E-VKK-..-E..H-.....---...-.-K-.NS
A_RW4.439cons IR----------N-L----K----D -S---------.----..-----..T---S.-..-.-.-----.--- ----S---E.--KEI----.---..--E....--...-.-S-.DK
A_RW4.538cons -K--C----------D-A---- ----------.--H-..-----..--------..-.-.-----.--- -----A--R.---K--M---.KS..LL.....--T..-.---.--
A_RW4.564cons -R--A----D-----M-H-NTS-Q-- ----------.----..-----..T-----E..-.-.-----.-Y- -L---A--Q.---K-V----..G-..--G....-T...-K-R-.T-
A_RW4.566cons I-V-F----D---S----FDKA-E-- ----------.GIH-..-----..----A-E..-.-.----K.-Y- T-NK---T-..---K--EK-..KR..H-K....--...-.---.--
A_RW4.618cons -R---------V---L----K----- ---------T.-I--..-----..--F---G..-.-.---K-.--- T-N-S---N.-----V----..--..--G....--...-.---.--
A_RW4.730cons -Q----------------FN-T-R-- ----------.--H-..-----..----I---..V.-.-P-KK.-Y- D--GI--ER.--KE--K---..G-..H-I....--...-.-N-.TK
A_TZ.TAN142 -R----------N----FA------ -I---------.--H-..-----..--------..-.-.-----.--- ----AP-RE.A
A_TZ.TAN15CON -K----------N-----T---E-- -I---------.--H-..-----..TS-----..-.-.-----.--- -----K--T.---K-VK---.KQ..--G....--...-.---.T-
A_TZ.TAN8 ----------N----FIK----- ----------.--H-..-----..--------..-.-.-----.--- ----A-----.-----VQ---..-T..H-G....--...-.-T-.-R
A_TZ.TAN9CON -R----------N----FSR----- ---------R.--H-..-----..--------..-.-.----R.--- ----A---T.--KE--D---..-I..--E....--...-.---.--
A_TZ2.016 ----------V-N----FT------ -S---------.----..-----..--------..-.-.-----.--- ----AR---.---K-VN---..-T..--G....--...-.---.-K-.A----.-------.---
A_TZ2.017 ----------V-N-----T---Q-- ----S-----.----..-----..--------..-.-.-----.--- ----AG---.---E--K---.ST..H-E....--...-.---.T-----.-------.----.S
A_TZ3.1574 -----V---L---TQ--E-- ----------.-I--..-----..----.--..-.-.-----.--- -----Q--N.--K--VD---..-T..--K....-A...-.---.PE-.-----.---S---.----.-------
A_TZ3.1576 -----HQEH----N---E-- ----------.--H-..-----..--------..-.-.-----.--- ----AK--E.A--K-V----..-T..--G....--...-.---.---.T-----.-------.----.-------
A_TZ3.1577 -------N-L--FAK----D ----F----T.--H-..-----..--F---A..-.-.----K.--- ----S---N.--L---D---..-..H-N....--...-.---.--P.-----.-------.----.G------
A_TZ3.1584 -----V-------D---E-- ----------.-IH-..-----..--------..-.-.-----.--- ---KSK--N.----------..-N..--G....--...-.---.---.A----.-------.--M-.-----D-
A_UG.1033 ----------N----FT---T-- ----------.-IHL..---R..--------..-.-.-----.--- ----AD---.--KE-VK-
A_UG.92UG037 ----------V-N-----N-S-T-- ---------R.----..-----..T-------..-.-.-----.--- ---GSQ---.--H--VE---..--..-WN....-N...-.---.NS-.-----.-------.--A-.-------
A_UG.964 -T--A--------N----FAR--R-- ---------E.----..-----..--------..-.-.-----.--- ---KE---N.---E--K---.Q-..ELGN...--...-.---.---.-----.-------.----G-------
A_UG.U455 See inserts
A_UG.UG06 -K------SD---------TK--T-- ----YKKV-R.RIH-..---R..S--TSN...L..------.-Y- ---GSQ---.-----V-K-...-E..-W.....-T...-.-N-.SKP.-----.-------.--V-.------
A_UG1.W2UG029 See inserts
A_UG1.W2UG03 --S----------HT--N-S -I---------.----..-----..--------..-.-.-N---.--- T-K-S-----.--T--VKE-..--..-W.....--...-.---.T--.-
A_UG1.W2UG037 -----V-N-----N-S-T-- ---------R.----..-----..T-------..-.-.-----.--- ---GSQ---.--H--VE---..--..-WN....-N...-.---.NS-.-
A_UG2.11 -------L-----N-----TK----- ---------T.--H-..-----..----R--..-.-.-----.--- ---KV----.-----VS---..--
A_UG2.115 ----------N-----T---I-- ----------.-IH-..-----..----R---..-.-.-N---.--- ---K.A---.-----VS---..-T..--.....-T...-.---.---.T---P.-
A_UG2.116 ----------N-L--FNTS-Q-- ----------.--H-..-----..------E..-.-.-----.--- ---G-A---.---E--K---..-E..SLGN...--...-.---.DK-.----P.--
A_UG2.117 I----------N-L--FN-T-Q-- ----------.----..-----..T-------..-.-.-N---.--- -I-GAK---.------KK-..GD..LL.....--T..-.---.NT-.--
A_UG2.118 -R---K-------N----FKT----- ----------.G-H-..-----..--------..-.-.-N--R.--- E-N-----R.---E--K---..-T..S-G....--...-.---.---.T---P.--
A_UG2.119 II-----L-----N-----KN--I-- ---------Q.--H-..---K..-------G..-.V.-----.-Y- ----AA---.---G--L---..--..H-E....--...-.---.T--.----P.--
A_UG2.124 -SD--------FN-S-T-- ----------.--H-..-----..-------A..-.-.-S---.--- -----K-RE.A--K--D---.E-..H-K....--...-.-
A_UG2.72 ----------N----FTK----- ---------R.--H-..-----..--------..-.-.-----.--- ----K---N.---E-VK---..-T..--G
A_UG2.78 -K-K----SD----L----TK----- ----G----R.----..-----..T-------..-.-.-----.--- ----S--KE.---K-VK---..-T..HW.....--...-.---.T--.-
A_UG2.82 -V-----F-----N----FTK--N-T -I--------T.----..-----..--------..-.-.----K.-Y- E-NK----R.---E-V----..-T..E-R....--...-.---.---.T---P.--

```
                                                        ______________ V3 LOOP _________________
                 ^^^                        ^  *^    ^^^                                         * ^^^     ^^ ^                             ^^   ^      ^^ ^              .  *       *
A_CONSENSUS_96   VMIRSENITNNAKTIIVQLVEPVKIN CTRPNNNTRK.SVRI..GPGQ..AFYATGD..I.I.GDIRQ.AHC NVSRTEWNK.TLQQVATQL..RK..YF.....NK...T.IIF.ANS.SGGDL.EITTHSF.NCGG.EFFYCNT
A_UG2.84         I-----------N----FT------- ----------.-IG-..----..T---ADN..-.-.-----.--- ---G-K---.---E--K--..-E..H-GR...--...-.---.---.----P.--
A_UG4.UG273      -K-K----SD----L----TT----- ----G----T.----..----..-------..-.-.-----.--- ----S--KE.---K-VK--..-T..HW.....--...-.---.T--.-----.-------.----.------
A_UG4.UG275      -K-----------N-L-.-TT--T-- ---------T.----..----..S------..-.-.---K-.--- -I--A---E.---K-VS--..-T..H-G....--...-.---.G--.----I.-------.---R.-----D
A_UG5.94UG003                          -E-- -V--------.G-H-..----..-------..-.-.-----.-Y- E-NG-K--E.---E--N--..KT..--R....-N...-.---.T--.-----.---PP--.----.A-LL
A_UG5.94UG004                          ---I ---------T.-IH-..----..T--T---..-.-.-----.-Y- T-NG----D.------E--..TT..N-V....--...-.-S-.-K-.----I.-T-----.----.---L
A_UG5.94UG009                          -E-S ----------.-I--..----..-------..-.-.-N---.--- ----AK--E.--KR--N--..-T..H-G....--...-.---.GR-.L---I.-------.----.----
A_UG5.94UG011                          -E-- ----------.--HL..----..------A..-.-.-----.--- -G-KAK--H.--YK-----..TP..--E....--...-.---.---.----V.------L.IV--.--LL
A_UG5.94UG012                          ---- ----------.----..----..-------..-.-.----K.--- ---SSS--R.--EXDKQK-..AE..HYP....--...-.-A-.NS-.-----.-V-----.--R-.----
A_UG5.94UG016                          -T-- ----G-----.-IH-..----..-------..-.-.-----.--- K---S-QHG.--G--VE--..-T..--N....-N...-.-T-.---.----I.-----G-.----.---L
A_UG5.94UG018                          ---- ----------.G-H-..----..-------..-.-.-----.--- -------H-.A-EKI-K--..GT..--......--T.D-.-VL.---.A----.-------.----.----
A_UG5.94UG019                          -T-- ----T-----.-I--..----..-------..-.-.----K.--- ---SSS--R.--EHG-K--..-E..HYN....-N...-.---.NS-.-----.--V----.----.----
A_ZR.Z32         -R-----F-D---I------K--N-T -M---------.-IS-..---R..--F----..-.-.-----.--- --------D.--SK--A--..--..H-V....-TS..-D---.---.----V.-------.----.------
A_ZR1.6557.sh        ----S-----------N--E-- ----S--R-Q.----..----..-------..-.-.----R.--- T-N--K--N.--D--KA--..E-..--N....-T
A_ZR1.6559           ---------------D------ ---------SQ.G-H-..----..V---RDR..-.-.----K.-Y- E-N-AV--N.---R--A--..GS..--G....--...-.-N-.TR-.-----.-------.--A-.------
A_ZR1.6563           ---F-D-G-----------N-S ----------.----..----..T------..-.-.-----.-Y- TIN------.-------R-..-E..--TT...-TT..-.---.--H.----
A_ZR1.6569       -A-------------------KQ-R-- -I-------R.GIHM..----..IL----S..-.-.-----.--- ---KKD--E.A-HK-----..-N..H-N....--...-.---.NS-.-----.---
A_ZR1.6571.sh        -A---D---N----FN-S-Q-- ----------.-I--..----..V--TN.-..-.-.-----.--- -I---A-YA.---K-VK--..--..H-P....-
A_ZR1.6649           ----------------T--P-- ----------.--H-..----..-------..-.-.-N---.--- ----A--S-.---D-----..-T..H-N....--...-.---.NRP.---
A_ZR1.6653           ---F------L----DS----- ----------.-I--..----..-------..-.-.-N---.--- ----AA--N.---K---K-..--..--G....-R...-.---.NS-.-----.-------
A_ZR1.6655.sh        ---------------N-S-S-- ----------.-I--..----..V-----
A_ZR1.6657           ---------------F-N--Q-- ----------.----..----..T------..-.-.-----.--- ----E---E.--HK-VE--..GT..--R....-T...-N-T-.-S-.-----.-------.----.------
A_ZR1.6663           ---F-----------NKT-Q-- ---------T.----..----..-------..-.-.----R.--- ----AA--N.--HK--VE-..HT..L-A....-T...-.VV-.TS-.V--
A_ZR3.K114       -R-----F-D-T-N-----AK--T-- ---------R.-IH-..A---..------A..-.-.-----.--- ---KKD--D.--Y------..GE..Q-GA...-I...-...-.-EP.----V.-------.----.-------
A_ZR3.L414       IA------S---------F-K--E-- -I-------R.----..----..T------..-.-.----K.-Y- ----KA--D.--L---K--..Q-..--E....--...-.-A-.NK-.-----.-------.----.-------
A_ZR3.P104        V-----------------DK--I-- ----S----R.GIHL..----..-----NA..-.-.-----.--- ---G-K--E.--HK--V--..--..H-GE...--...-.-V-.TKP.----V.-------.----.-------
A_ZR3.PZ61        VV----------------D---I-- ----------.GIHL..----..T-C---E..-.-.-----.--- ---G----E.A--K-VG--..--..H-GE...--...-.-V-.T-P.----V.-------.----.-------
```

```
                                                  V3 LOOP
                  ^^^         ^^^   ^ **  ^^^                    * ^^^   ^^ ^                ^^   ^    ^^^                *      *
B_CONSENSUS_96  VVIRSENFTDNAKTIIVQLNESVEIN CTRPNNNTRK.SIHI..GPGR..AFYTTGE..I.I.GDIRQ.AHC NLSRAKWNN.TLKQIVKKL..RE..QFG....NK...T.IVF.NQS.SGGDP.EIVMHSF.NCGG.EFFYCNT
B_AR.21281       -------X-------X----T---- ----------.--TL..-----..VY-----..-.X.-----.--- -I-X-----.------X--..-..-.---....--..-
B_AR1.06        I--------N----------X-- --------X-.---M..AG--..-I-A--R..-.-.----R.SY- ---STE-TK.--QGVA----..N-..--.....-..-.---.-H-.-----.--------.I---
B_AR1.13        I--------N----------V-- ----S-----.G---..----..-------..-.-.-----.--- ----TE-GK.-----EV---..K-..--N....--...-.-Y-.---.-----.--------.----
B_AR1.21        I-----S------N--I------V-D -I---------.--NM..-A---..-W-A----..-.-.----K.--- --TGT--E-..--R---AS---..K-..-.-.....--
B_AR1.22        I--------------------IV-- ----S---S-.---M..AG--..-L-A--R..-.-.-----.--- ----T--EE.-----G----..K-..K-.....-..-.-Y-.---.-----.--------.-----.--
B_AR1.23        -I---D--S----------T---- ----------.G--F..A---..TP-A--D..-.-.-----.--- -V--------.--RL--E---..-K..EYK....-AT..-.-A-.-H-.-----.---L---.---
B_AR1.24        I--------N-V---------T---- ----G-----.---L..-F---.TL-A--N..-.-.-----.--- -I-E----K.-----AM---..K-..-.-.....-..-.-A-.---.-----.--------.----
B_AR1.29        I--------N----------V-- ----------.G--M..-W---..---A----..-.-.----K.--- -I--KH-ID.---KVGV---..K-..-.-.....-..-.-I-.KH-.-----.-----T-.----
B_AR1.33        -----A--S----------K----- -X--------.--PM..---K..---A---..-.-.-----.--- --------X.--R---T---..KG..---.....-..-.-Y-.---.---G-.--I----.---
B_AR1.34        I--------N----------V-- ----S---S-.---M..------..------R..-.-.----K.--- ---STQ-MD.---K-GI---..-.-..--N....--...S.VI-.TH-.---G-.-------.----
B_AR1.35        I------------------K--A-- --------S-.---M..AW---..T--A--R..-.-.-----.--- ---KTE-RQ.-----ET---..-.-..--N....--...-.-I-.SN-.------.--E----.----
B_AR1.56        I----K-------I---------V-- ----S---S-.---L..AG-K..-L-A--R..-.-.-----.--- ----TQ-EK.--E--AI---..-.-..--N....--...-.-I-.KN-.------.--T----.---
B_AU1.MRC       --------N---------KKP---- ----------.D---..-----..-IFR---..-.-.-----.--- -V-GT---D.------I---..-.-..--K....--...-.---.R-.------.-------.----.-------
B_AU1.MRC2      --------N---------K-P---S ----------.--P-..-----..--W-----..-.-.----.--- KVNET--KD.--R--AE---..-.-..--K....--...-.-I-.---.-----.--E----.-----.K-----S
B_AU1.MRC3      -----D-------------KDP-N-T -V-------R.-----..----..---A--D..-.-.-----.--- ----ED-HK.A-E--AG---..-.-..--N....--...-.---.-R-.----L.-V-V-T-.-----.-L-----
B_AU2.C18C      See inserts
B_BE.SIMI84     See inserts
B_BR.002con       II-------N----------S--V-- ----------.--RM..-L--W.SV-A--K..-.-.-----.--- ----RE-EK.-----AV---..Q-..-.-.....-N...-.-I-.KK-.------.--------.----.-------
B_BR.BZ           -I-----L---T-N-----K-A---- ----------.---M..-W---..---A--D..-.-.-----.--- ---------.------M---..-.-..H-.....--...-.-I-.---.-----.----T-.----.-----R-
B_BR1.10553           -----N---I-----KD--T-- ----------.--Q-..-----..-------..-.-.-----.--- .----D--K.--E----N--.G-..-.-.....H-....-.-A-.-R-.------.---T---.----
B_BR1.1056            -----K-------------V-- ----------.G--M..---S..T--A---..-.-.-----.--- ----T--E-..--R---I-I..--..--R....--...-.---.---.----S.--------.----.------
B_BR1.10565           ----------------Q----- ----S-----.G--L..-F---..-L-A----..-.T.-----.--- ---ST--EQ.-----VAR---..HD..-.-.....--...-.-F-.---.------.-------V
B_BR1.10575           ----------------K--V-- ----------.-LSL..-----..------D..-.V.-----.--- -----E-T-..--N----E-..-.-..-.....--...-.---.-H-.------.---R---.--R-.------
B_BR1.1058            -----N-------H--D----- ----S-----.G---..-----..T-FA---..-.-.-----.--- -I-E-----.--R-VAI---..-.-..-.....-G...-.-H-.---.------.-----T-.--A-.------
B_BR1.1059            -------V-V------D-L--- ----------.S-.G--M..---G..------R..-.-.----R.-Y- -I--------.-----AR---..G-..--V....--...-.---.---.------.---T---.----.--S--
B_BR1.10593           ---I----------KDP----- ----------.--S-..-----..---A--D..-.-.-----.--- -I------D.--Q-------..-Q..--K....--...-.---.---.----Q.-V-----.-----.------
B_BR1.7930            -----N-------H-------- ----------.G--M..-W---..T--A---..-.-.-A---.P-- -I--E---K.--E--A---..Q-..-N....IT...-.-H-.-R-.A----.--R--T-.-----.-------
B_BR1.7932            -----N-V-------------- ----------.---M..-W---..---A--D..-.-.-----.--- -------EK.-----AG---..AK..LYN....--...-.-I-.---.------.---T---.----.------
B_BR1.7934            -----N-------------V-- ----------.---V..-W---..SLF-----..-.-.-N--L.--- ----I--E-..--N--AI---..K-..--V....--...-.-R-.------.---T---.----.------
B_BR1.7936              L-N----------T------ ----------.---M..-W---..---A---..-.-.----E.--- -I--------.----------..K-..-.-.....--...-.-I-.-H-.------
B_BR1.7940            ---------S---H--V----- ----------.R-Y-..-----..-V----Q..-.-.----R.--- -I--------.--R---I---..-.-..--K....R-...-.---.KP-.------.---T---.----.------..
B_BR1.7942            ----SN----------KT-N-D -E---------.--N-..-----..------D..-.-.-----.--- -V-KND-L-.A----AI---..G-..-.-.....--...-.---.---.------.--------.--R-.---F-
B_BR1.7946            -----N----------T-K-- ----S-----.---L..-L---..------D..-.-.-----.--- -I---E---.--R--AQ---..-K..--E....--...-.---.-H-.------.-----T-.-----
B_BR1.8615            -------T-I-----KDH---- ----H----R.--T-..-----..------D..-.-.-----.--- -V--VQ---.--A---Q---..-.-..-D....--...-.---.---.------.------T-
B_BR1.862             -------T---------T---- ----S-----.---L..-W---..---A---..-.-.-----.--- ---T---E-..--E--A---..ED..-.-.....--...-.---.---.------.--------.----.-------
B_BR1.8623            -----N----------TT-K-- ---L-----T.----..---Q..---A--D..-.-.-----.--- -----D--D.------T---..G-..---S...-T...I.TI-.TS-.------
B_BR1.8625            Y----N-----------A--- ----------.-----..----S..---A--D..-.-.-----.--- -----E-D-..--T-V----..A-..-.-.....--...-.-K-.-E-.--------.I
B_BR1.8629            ---L-----------T-I--- ----------.---M..-W---..T--A----..-.-.-----.--- -I-TE---T.--E--A---..VD..-.-.....--...-.-I-.---.------.-VI--T-.----.------
B_BR1.8633            YG-L-G-T-----K-KDN---- ----------.G---..----..---A-.-..-.T.-----.--- -V--T-----.--A-------..-.-..--P....--...-.---.---.AE---.--------.--
B_BR1.8635            ---I----------KT---- -L--S-----.---M..-W---..---A---..-.-.-----.--- ---K---EE.--N-VAI---..-.-..--EN..N-...I.T--.K--.------.--------.----.------
B_BR2.W2BR003           -I---K---------T---- ----------.---M..-W---..---A---..-.-.-N---.--- ----------.--R---I---..K-..H-.....--...-.-I-.---.-
B_BR2.W2BR004           -L---T-I---------V-- S---G-----.G-P-..---G..S--A-ER..-.-.-----.--- X---T--E-.----VAD---..-.-..--K....--...-.---.---.-
B_BR2.W2BR014           --MN----------N----- -I--------R.--I-..-----..---A--D..-.-.-----.-Y- T-NSE----.------T---..-.-..K-....--...-.---.-HA.-
B_BR2.W2BR017           ---N---------K-A-Q-- ----------.----..-----..---A--D..-.-.-----.--- -I---Q---.------E---..-K..--D....--...-.-A-.KP-.-
B_BR2.W2BR018           ---N----------T-Q-- --------X-.----..---S..---A--D..-.-.-----.--- -----Q---.---------..-.-..-K....-S...-.-A-.KP-.-
B_BR2.W2BR019           ---N----------T----- ----G----R.---M..-W---..---A--D..-.-.-----.--- -------E-.------ATX-..-G..--P....--...-.---.---.-
B_BR2.W2BR02            -------I-------T---- ----------.---M..-W---..---A---..-.-.-N---.--- ---I---S-.-----ER---..KA..-.-.....-N...-.---.-H-.-
B_BR2.W2BR020           -L-N---------KDP-D-- ----------.----..----..---A--D..-.-.-----.--- -----Q--D.--SK--T---..-.-..--E....--...-.-K-.QPP.-
B_BR2.W2BR023           ---N----------------- ----------.-----..----K..---A---..-.-.-N---.-Y- -I--TA-RK.--REVSR---..A-..H-P....--...-.-I-.---.-
B_BR2.W2BR024           -----Y---------TI--- ----------.---L..-W---..---A---..-.V.-----.--- ---KDI-EK.---R-AI---..G-..--K....--...-.-A-.-K-.-
B_BR2.W2BR026           ---S-T---------P---- ----------.--T-..----..---A---..-.-.-----.--- I-NKT--ES.--E----N-..Q-..-.-.....--...-.---.---.T
B_BR2.W2BR028           -L-N----------T-I--- ----------.---M..-W---..T--A---..-.-.-----.--- -I-SN--RT.--G--A----..EX...      -X...I.-QV
B_BR2.W2BR030           ---------------P-Q-- ----S-----.G---..----..---A--D..-.-.-----.--- -I---Q---.------G---..-.-..---....--...-.---.---.-
B_BR3.HRJ17     ---------N----------V-- ----------.---M..----..-Y-----..-.L.-----.-A-
B_BR3.HRJ27.sh                          ---------R.-----..-----..---GG-I..N.K.EN---.---
B_BR3.HRJ477     ------I-N-V----------V-- ---------R.---M..-L---..---A.-D..-.-.-----.-Y
B_BR3.HRJ625    ----------V---------T---- ----------.----..----..-WFAR-G..T.T.-----.--- ---ST----.------I---..Q-..---....--...A.-I-.K--.--RG-.RN
B_BR3.HRJ626    ---------N----------IA--- ----------.--R-..---S.---A--D..-.-.-----.--- ----T--EH.--T--GR---..Q-..-N....--...-.-I-.---.-----.--------.----.-------
B_BR3.HRJ636     G-------N---I------Q--V-- ----------.--P-..A--S..-WFA---..-.-.-----.--- ---KVH-E-.-----A---..-.-..--E....--...-.-Q-.---.-----.--
B_BR3.HRJ70                      YVK-S- ----S-----.---L..-W---..-------..-.-.-----.---
```

V3 LOOP
B_CONSENSUS_96 VVIRSENFTDNAKTIIVQLNESVEIN CTRPNNNTRK.SIHI..GPGR..AFYTTGE..I.I.GDIRQ.AHC NLSRAKWNN.TLKQIVKKL..RE..QFG....NK...T.IVF.NQS.SGGDP.EIVMHSF.NCGG.EFFYCNT
B_BR3.HRJI01 II---Q-IS----I---H-------- ----------.--N-..----..---A--D..-.-.------.-Y- -I---Q---..--E--AI--..G-..--K....--...K.-A-.T--.-----
B_BR3.HRJI04 -I---H-------- ----------.G---..-F--..TLFA-A-..-.-.-----.--- -I--------.A-TK-AM--..K-..--.....-S...-.---.K--.-RRG
B_BR3.HSP204 See inserts
B_BR3.HSP205 -I-----I----------V-- ----S-----.----..----..---A---..-.-.-----.--- ---G-A-D-.------T--..--..--.....--...-.---.---.------.N
B_BR3.HSP228A2 YV--N----------KT-A-- ----------.---M..-W--..---A---..-.-.-K---.--- -----E---.-----AIR-..EN..--.. -.NIS.YH-.------.-------.-----.--------
B_BR3.RJ12 I------L-N----------K--V-- ----S-----.-----..-W--..-I-A--A..-.-.-----.--- ----K--E-.--E--A---..--..--.....--...-.---.---.------.---
B_BR3.RJ14 --V------N---I---H---T---- ----------.---V..-W--..-L-----..-.-.-N---.--- -I-GT--G-.---W-AA--..K-..--.....--...-.-K-.QPP..QEET
B_BR3.RJ19 G-----D--N---I---------- ----------.--QY..-T-GG.---A---..-.V.-----.--- -I-KE----.-----AT--..K-..--.....--..-.---.---.---
B_BR3.RJ379 -----D-------I---------V-- ----G-----.----..----..-------D..-.-.-----.--- --------D.--R--AI--..--..--N....GTE..-.-I-.---.---E-
B_BR3.RJ478 ---------N----------K----- ------Q---.---M..-W--..--H-N--..-.-.-N---.--- ---GIE---.--T--AE--..--..---.....--...-.-I-.T-P.----
B_BR3.RJ483 I--------N---V---------- ----------.G--M..-L--..---A--G..-.V.-----.--- -I-----K-.--R--A----..K-..--N....--...-.---.---.------.-------.----.-------
B_BR3.RJ484 --L--------V---------- ----S-----.G---..-W--..---A---..-.T.----K.-Y- -I--T--E-.--R--AI--..E-..--K....--...-.-I-.---.------.---T---.-----.------D
B_BR3.RJ49 -------I-N----------V-- -S---------.---M..-W--..-------..-.-.-----.--- -V--ED-K-.--EK-AR--..--..H-K....E-...P.---.-KT.----
B_BR3.RJ54 -------I----------V-- ----------.----..----..-------..-.-.-----.--- -----Q---.---L--T--..--..---.....--...-.-A-.-S-.-----
B_BR3.RJ59 ---------N---------- ----G-----.---L..-W--..---A--A..-.-.-----.--- T-NKT--EK.--E---I--..--..--V....--...-.-T-.---.----
B_BR3.RJ62 --V------N---I---------- ----S-----.---L..-W--..---A---..-.V.----E.--- ----T-----.-----AI--..K-..----.....--...-.-E-.K-N.----
B_BR3.RJ623 --V----------I-------T---- ----S----R.---L..---G..-------..-.-.-N--K.-F- T-NGT--E-.-----AE--..--..--.....--...-.-A-.K-N.-----
B_BR3.RJ626 ---------N----------IA-- ----------.--R-..---S..---A--D..-.-.-----.--- ----T--EH.--T--GR---..Q-..--N....--...-.-I-.---.-----.-------.----.------
B_BR3.RJ636 G.--------N---I------Q--V-- ----------.--P-..A--S..-WFA---..-.-.-----.--- ---KVH-E-.-----A---..--..--E....--...-.-Q-.---.------.--
B_BR3.RJ64 ---------N---------- ----------.---L..-W--..------Q..-.-.-E---.--- -------EK.-----AI--..--..---.....--...-.VI-.---.---G-.......R.--N
B_BR3.RJI02 I-V------N---I---------- ----------.---V..-V--..-I-A---..-.-.-----.--- -I-ETR-K-.--EK-A---..K-..-VV....--...K.-S-.Q-A.------.-
B_BR3.SPB I--------N-------H-----V-- ----S-----.---M..-W--..---A---..-.-.----R.--- -----A--S.--R--R
B_BR3.SPB2 I----K-----T-I-------T---- ----------.---M..-W--..------D..-.-.-----.--- -I-
B_BR3.SPB3 I--K----S----------T-K-D -I---------.G--M..-W---..T--A--R..-.-.-----.--- ---KVA--R.--ER-AI--..-N..--NY...-N...D.KN-.---.-----.-------.----.-----D-
B_BR3.SPB4 -G----------M-----TP---Y ----------.RKS-..-----..-------..V.-.-----.--- ----KR--D.--I---TE-..--..H-K....--...-.---.-----.---I-C-.---R
B_BR4.BZ167A ----------L---K-P---- ------KA-R.R-R-..-----..T--.--K..-.V.-----.-Y- DI--------.--N---T--..--IE--K....--...-.---.---.-----.-------.--R-.------S
B_BR4.BZ200A ---------N----------G--V-- ----S-----.---L..-F--..-L-A---..-.-.-----.--- I-NGTE--K.--N--AI--..--..---G...--...-.---.---.-----.-------.----.-------
B_BR5.504 -L-N-V-----H-K-H---- ----G-----.G---..---K..----R--..-.-.-----.--- -I------T.--Q-VAI--..--..---.....-T...-.---.--K-.----
B_BR5.505 -I-N----------K ----F-T--R.---T..-W--..-LF---..-.-.-----.--- ---G---ED.------I---.G-..--K....--...-.-A-.-R-.-----
B_BR5.506 ----------V-- ----------.--QL..-M--..-L-A---..-.-.-----.--- -I--EN-E-.------I---.G-..--.....--T..-.TI-.F--.--
B_BR5.507 -I--S----------T----- ----S-----.G--M..-W--..-T--A--Q..-.-.-----.--- ---SV-----.------R---.K-..H-.....--...-.-I-.---.-----
B_BR5.51 ---N---------T---V-- -----T----.---M..-V--..-L-A--D..-.-.-----.--- ---KVQ-EK.-----AI---..-A..--N....--...-.---.---.-----
B_BR5.510 ---V-T----------Q-- ----------.--P-..----..------A..-.-.-----.--- -----Q---.--E--AT--..--..--K....--...-.---.---.-----
B_BR5.512 ---N---I------K--V-- ----------.---M..-W--..----N--..-.-.-----.--- T---LD-E-.----VA---..--..--N....--...-.---.---.-----
B_BR5.513 -----T----I--K-H---- ----------.G---..-----..T--A----..-.-.-----.--- -V-G-----.--L------..--..-YN....IR...-.---.---.-----
B_BR5.514 ---N----------T---- ----------.---M..-W--..T--A---..-.-.---..T--- ---------.-----AI---.E-..--.....--...-.-I-.---.-----
B_BR5.515 --SN--N--------T---- ----------.G---..-----..V--A---..-.-.---.-.--- -I-S-----.--TK--H---..--..---.....--...-.--.-H-.-----
B_BR6.P3 I------I-N----------T-V-- ----------.---M..-W--..T--A----..-.-.-----.--- ---K---E-.--N--AI--..K-..--.....-N...-.-S-.---.------.--I----.-----.-------
B_BR6.P4 I----------------R---V-- -S---------.RMTL..-----..VY------..-.V.----K.--- -I--T-----.--EK--G--..--..---.....--...-.--.---.-----.-----.------S
B_BR6.P6 I--------N----------T---- ----------.G--M..-M---..-L-V----..-.-.------.-Y- ----T--S-.-----AI---..K-..--.....--...-.-F-.---.-----.-----T-.----.-------
B_BR6.P7 -----D---N----------V------ ----S-KK-R.R---..-----..--IG-..-.-.T.-----.--- -M--------.----V-T--..--..---.....--...-.---.---.-----.-----T-.----.---F---
B_BR6.P8 ---------N-------H--K--A-- ----S-----.---V..-F--..-L-A----..-.-.-----.--- ---KTE-ER.--E-V-I--..--..---S...--...-.-N-.T--.-----.-----T-.----.------S
B_BR6.P9 I--------N----------KP---- ----------.G---..-----..-V---RK..-.-.-----.--- -I---L--D.------T--..--..-V-.....-S...-.-D-.KP-.----L.---T---.-WE-.-------
B_CH1.K1 ---------N---------KQP-V-- ----------.----..----..-------..-.-.-N---.--- ----SA---.--
B_CH1.K12 ----------------ID----- -------K--.----..A---..---A--V..-.-.-----.--- ----T--ED.--
B_CH1.K23 ---------N---I--------IK-E ---------R.-V-V..---K..-L---.H..-.-.-----.--- -I---N---.--
B_CH1.K26 --------RN----------T-Q-- -I--S-----.----..----..---A--D..-.-.-----.--- -I-G-Q--K.--
B_CH1.K28 ---------N-V----------A-- ----------.--S-..-----..---A--D..-.-.----K.--- -----Q---.--
B_CH1.K32 ----------I-----K---Q-- ----S-----.--P-..---K..VL-----..-.-.-----.--- ---------.--
B_CH1.K47 ---------N-V---------- --------S-.G---..-----..T--A--R..-.-.-----.--- -I-IT----.--
B_CH1.K53 ------T----------RK----- ----------.----..----..---A-.D..-.-.-N---.--- -I-GVE---.--
B_CH1.K6 -----A---N--RI---------A-- ----------.--P-..---K..--------..-.-.-----.--- ---K-E---.--
B_CH1.K65 ---------N----------IA-E -I---------.-----.---K..-L---.G..V.-.------.-Y- --N-T---D.--
B_CH1.K72 -----D-L----------N- ----------.--SM..---K..-I----A..-.-.-N---.-Y- -I-IT----.--
B_CH1.K8 ------DLSN---------- ---------R.---M..---K..-I----A..-.-.----R.-Y- -I-IT----.--
B_CH1.K9 ---------- ----------.-----..----..---AS--..-.-.-----.--- ---E-----.--
B_CH1.P --------RN---------- ----------.-----..----..--------..-.-.------.-Y- -V-I----R.--
B_CH1.P2 -I-------N----------T-Q-- ----------.--N-..----..--------..-.-.-----.--- -I-----E-.--
B_CH1.P3 -----D-I-N--R------TDT-V-- ----------.----..----..---A----..-.-.-----.--- -I-------.--
B_CH1.P5 ------K----------K--IKN- ----------.-----..----..---A--N..-.-.-----.--- -----D---.--
B_CH1.P6 -I-----L---------R---T---- ----------.--TM..-----..T--A---..V.-.-N---.--- --N------.--

```
                                                                V3 LOOP
                 ^^^           ^^^   ^ *^    ^^^                                  * ^^^    ^^ ^                      ^^   ^     ^^^                 *       *
B_CONSENSUS_96   VVIRSENFTDNAKTIIVQLNESVEIN CTRPNNNTRK.SIHI..GPGR..AFYTTGE..I.I.GDIRQ.AHC NLSRAKWNN.TLKQIVKKL..RE..QFG....NK...T.IVF.NQS.SGGDP.EIVMHSF.NCGG.EFFYCNT
B_CY1.HO21       ---------NY----------V-- ----------K--N-..E---..---A-A-..-.-.-----.--- ---SKA---.--R--AI--..K-..---E...--...-.-S-.---
B_CY1.HO25       ---------N---------K-P---- ----G---I-.G-YM..-----..--R--KK..-.-.---K-.--S -I-TQ-----.----V-I--..-K..----...--...-.-I-.K-P
B_CY1.HO27       I--------N-------------I-- ----------.--N-..----..---AIE-..-.-.V---A.--- ---G-V---.--RRL-I--..--..P--....Y-...-.-I-.--P
B_CY1.HO28       -------------------K------ -------I--.----..------..-.-.-----.--- -------D-.------R--.--..---.....--...-.---.---
B_CY1.HO294      ---------N-------------V-- -I--------.--N-..----..---A--Q..-.-.----R.--- ---SKA---.--RK--I--..--..---K...--...S.-S-.---
B_CY1.HO39       ---------N-----------TI--- ----------.----..----..--FA--D..-.-.-----.--- -I-KT----.--Y--AT--..--..-YV....--...-.-S-.-H-
B_CY1.HO40       I----A--SN-V-------K-A---- ----------.R-TM..-----..VW-----..-T.----K.-Y- T-NGT--H-.----VAA--..--..--.....--...-.-I-.---
B_CY1.HO433      -I---D---N---I------KT---- ---------R.-----..----..---A--D..-.-.----L.-Y- -I-G-E-K-.------G--LEI-..--K....--...-.---.---
B_CY1.HO45       ----------------------T-K-T ----------.G---..----..T--A---..-.-.-----.--- -----Q---.------I--..--..--V....--...-.---.---
B_CY1.HO464      I----S-L---T-I------------ ----------.--PM..----..---A--D..V.-.-----.--- -I-GE----.--RK-AA--..K-..--.....--...-.---.---
B_CY1.HO48       ---------N--------H-------- ----S-----.G---..----..-----EQ..-.V.----K.--- -I-A------.--N---E--..G-..--K....--...-.-I-.QP-
B_CY1.HO503      I-------S-----------K------ ---------R.--N-..----..---A-.D..-.-.-----.--- ----QD--D.------T--..--..--K....--...-.---.---
B_CZ.BTSPR       -------------N-----K------ ----------.---M..----..---A--D..-.-.-----.--- -I-K-----.------T--..--..--K....--...-.---.---
B_DE.D3          -----D-------------K------ ------Y-S-.R-R-..-AR-..----K-K..-.-.-----.--- -I-G---DS.--R------..--..R---....--...-.---.---.-----.---T---.----.------
B_DE.HAN         -----D-----T-----H-------- ----------.G---..----..-V----R..-.V.----L.--- -I---R--K.--N--FR--..--IR--E....--...-.---.-R-.-----.-------.----.------
B_DE.serocons    ---------N---I---------V-- --------R.--N-..----..---A----..-.-.-N--L.--- ----T--E-.--R-VAR--..--..--.....--...-.---
B_DE1.A                  -----I-----KKP---- -I--------.D---..----..-I-A--R..-.V.----K.--- -
B_DE1.C                   ----I-----K-P---- -----HXX-R.R---..----..---A-.N..-.V.----K.-Y- -
B_DE1.D                    ---I---------X-- ----------.--P-..----..-------..-.-.-----.--- -
B_DE1.E                  ---T-N-----X---V-- ----------.G---..---X..-------..-.-.-----.--- -
B_DE1.F                  -E-T-------------- ----S-----.G---..----..------Q..-.T.-----.--- -
B_DE1.I                  ----------KA-XX--- ----------.G---..----..---A---..-.-.-----.--- -
B_DE1.J                  ---------H-------- ----------.--P-..--EK..V-------..-.-.-----.--- -
B_DE2.3493       -----Q-------I------------ ----------.-----.A---..---A---..-.-.-----.--- -I------K.--EK--I--..--..---....--...-.---.-R
B_DE2.3497       -----D--SX---------T-P-V-- ------Y--X.G-P-..-----..---XX-D..-.-.-----.--- -I---E--X.--YXV-E--..-A..X-X....--...-.-A-.--
B_DE2.3498       ---------N---V---------V-- ----------.--N-..----..---A--D..-.-.-----.--- ---G-A---.--R---I---..KD..---E...--...-.---.QN
B_DE2.3499       -M-X-----N-VRN---------G-- ----X---S-.G---..----..---A-EA..-.T.-----.--- -I--------.--Q------..G-..--V....--...-.---.--
B_ES.03          -------L-----------KDP-Q-- ---------R.--T-..----..------D..-.-.------.-Y-
B_ES.06          -I----D----------------V-- -S---------.--P-..----..-------..-.-.-----.---
B_ES.07          -----D--S--T---------K-I--- ----------.--P-..----..---A--D..-.-.-----.---
B_ES.09          ---------------------T---- ----------.--P-..----..-------..-.-.-----.---
B_ES.10          I----D-----T---------AIK-- ----------.--P-..----..---A---..-.-.-----.---
B_ES.106         I------I---------H----I--- ----------.----..----..---A---..-.-.-N---.---
B_ES.108         I----------------H----IQ-- ----------.G---..---N..------.-..-.T.-----.---
B_ES.13          I-------------------T---- ---------R.--T-..---S..V--.---..-.-.-N---.---
B_ES.14          -----A-----------H-------- ---------R.----..---K..V--A.-D..-.-.-----.---
B_ES.15          -----D---N-V-------K-P---- ----------.--P-..----..-------..-.-.-----.---
B_ES.17          ---------NHI-------K-P-Q-- ---------R.----..---K..X--.--D..-.-.-N---.---
B_ES.18          --V-A--I---T-------K-AIQ-- ----------.--T-..----..-------..-.-.-----.---
B_ES.20          --V--A--S--T-------K-PI--- --------S-.G---..-----..T--A-.N..-.-.-----.---
B_ES.26          ---------NHV-------KDPIA-- ----------.G---..----..---A-.D..-.-.-----.---
B_ES.D17         -I--------N---------------- ----------.RVS-..----..-WR---Q..-.-.----K.-Y-
B_ES.III         -----V-------------TT-T-- ----------.G-P-..----..---A-AS..-.-.-V-T-.---
B_ES.IV          I--------N---------------- ----------.G---..----..-L---.-..-.-.-----.---
B_ES.IX          -------I---T---------TI--- ------Y---.RVS-..----..--R-A-S..-.-.-----.---
B_ES.P12         -----D-------S-------T---- -S---------.G--L..---G..------A..-.-.-----.---
B_ES.P13         II------S----------QTI--T -M--H----R.---L..---S..---A--D..-.-.-----.---
B_ES.P14         ----------V----------V-- -----KKVER.R---..----..---A-.D..-T-.-----.---
B_ES.P17         ----------------------P-K -I--------R.G-N-..----..---A-.D..-.-.-N--E.---
B_ES.P18         I----------L-N-------P-Q-K -I---------.G---..---K..---A--R..-.-.-----.---
B_ES.P19         -----D---N---------K-P---- -I---------.G---..----..V--A-.G..-.-.-----.---
B_ES.P20         -----D-------V------T-I--- -----IT---.---M..-----..TVHAI-R..-.-.-----.---
B_ES.P21         --------SN--R--------TIQ-- ----------.----..----..----S.-..-.-.-----.---
B_ES.P6          I----D-----V----------I--- ----S-----.----..----..-------..-.-.-----.---
B_ES.P9          --------KE-I-P---------- -I---------.-----..-----..--H---A..-.-.-----.---
B_ES.R1          ---------N---------K--IK-- ---------R.GV--..--R-..G----.-..-.-.-N---.---
B_ES.S1          -----D---N---------K-AI--- ----------.--P-..----..----I-D..-.-.--L--.---
B_ES.THF13A      I----A-------------A-I---- ---------R.G---..-----..---G-.D..-.T.-N---.---
B_ES.THF17       ---------N-------H----I--- ----------.----..---S..------R..-.-.-----.---
B_ES.V           -------------N-----K------ ----------.G---..----..---A--D..-.-.-----.---
```

V3 LOOP

```
                    ^^^          ^^^   ^ *^   ^^^                    • ^^^    ^^ ^                    ^^   ^     ^^^                  *       •
B_CONSENSUS_96  VVIRSENFTDNAKTIIVQLNESVEIN CTRPNNNTRK.SIHI..GPGR..AFYTTGE..I.I.GDIRQ.AHC NLSRAKWNN.TLKQIVKKL..RE..QFG....NK...T.IVF.NQS.SGGDP.EIVMHSF.NCGG.EFFYCNT
B_ES.VI         ----------V--------K----- ----------.--AL..---G..-V-A---..-.-.-----.---
B_ES.VIII       I--------N------------IA-- ----------.G--L..---G..--FA--N..-.-.-----.---
B_ES.X          ---------N---------K-P-K-- ----------.G---..----..------A..-.-.-----.---
B_FR.CA         -------I-N----V----K------ ---------R.R-T-..-----..VY----A..-.-.----K.--- -I--------.----------..--..--E....--...-.---.---.-----.-
B_FR.CB         ------KI-N---S-------T---- ----------.G---..----..---A--D..-.-.-N---.--- -I------D.--R--AI--..G-..--K....--...-.---.---.-----.-
B_FR.CC         I----D----------Q--V-- ----------.D---..----..---A----..-.-.-----.--- ----VQ---.--R-VAT--..--..--.....--K..-.V--.K--.-----.-
B_FR.CD         --------AN---------K-P-K-- ----------.--P-..---K..---A--D..-.-.-----.--- --------D.------N--..G-..--K....--...-.-S-..--.-----.-
B_FR.J6          ----D--S------------T---- ----------.----..----..---A---..-.-.-----.--- -I-------.------R--..--..-YE....--...-.---.K--.-----.-------.I---.------S
B_FR.J91         ----D--S------------T---- ----------.----..----..---A----..-.-.-----.--- -I-------.------R---.K-..-YK....--...-.---.K--.-----.-------.T---.------S
B_FR.LAI        -----A-------------Q----- ----------.--R-QR----..--V-I-K....-.-.NM--.--- -I------A.-----AS--..--..---N...--...-.-I-.K--.-----.---T---.------
B_GA.OYI        -I---S---N---I------K----- ---------N.R-S-..-----..--H--KQ..-.-.-----.--- -----T-EK.--E--AT--..-K..--R....--...-.-A-.DR-.-----.-------.----.------
B_GB.AIT        -----D-------I------K----- ----------.----..---K..TL-A----..-.-.-E---.-Y- ----TQ-KD.------I--..--..--K....--...-.---.---.-----.-....-.L-
B_GB.CAM        ---------N---------K-P---- ---L-------.--A-..-----..TV-A-DR..-.-.-----.--- ---ST----.------T--..K-..---....--...-.-I-.---.-----.-------.----.------
B_GB.Man        --L--D----------H-------- ----S--S---.--Y-..-----..R-HV-RA..V.T.-----.--- -I-K---EK.------E--..--..K--....--...-.-I-.---.-----.---T---.----.------S
B_GB.V12                  ----------N----- ----SK-I-R.-----..-S---..----IEG..V.A.--V-K.-Y- T-NGT---D.---L--A--..--..---....--...-.-Q-.KP-.-----.--
B_GB.V74            -----N-V----------V-- ---------R.G---..-----..---A--N..-.-.-----.--- -I--T---D.------T--..--..--E....--...-.-I-.---.-----.-------.-
B_GB.V77                ----------KDP-N-T ----S-----.----..-----..V-HA---..-.-.-----.--- -----D---.------T--..--..--E....--...-.-I-.---.-----.----
B_GB.V82            -S----------------V-- ----------.R-Y-..----..-V---EQ..-.-.-N---.--- -I------E.------I--..--..--E....--...-.----.K--.-----.-------.-
B_GB.V87             -------------KDP-K-- ----G----E.R-S-..-----..--IAR-Q..-.-.-----.--- -I-K-E---.--R---T--..K-..--K....--...-.-I-.---.-----.---
B_GB.V9                           K-- ----G-----.--P-..-----..--IA-SQ..-.-.----K.--- ----EE---.------T--..--..--K....--...-.-I-.-
B_GB1.CPHL    See inserts
B_GB2.CPHL2     ----------V------V------K -I---------.----..-----..------Q..-.-.-----.--- -I-EE--K-.--QRVA---..--..K-E....--...-.-A-.T-P.-----.--A----.----.-------
B_GB3.CPHL7   See inserts
B_GB4.AC03      ---------N----------T---S ----------.GV-L..---S..-W-A--G..-.-.-----.--- -I--V----.--EK--I--..-K..---....--...-.---.KP-.-----.---T---.--A-.---F-KS
B_GB4.JB14      -----D--S---N-------K----- -I---------.G-PM..-----..------Q..-.-.-----.--- -I---Q---.------I--..--..--N....-S...-.---.---.-----.-------.----.-------
B_GB4.M2347015  I--------SN---------TKP-V-- ----------.--P-..-----..---A--D..-.-.-----.--- ---KTD---.------E--..--..E--....--...-.-A-.---.-----.--E----.----.-------
B_GB4.M26864-01 ----------------H-------- ------K-KR.-F--..-----..--L.-.-..-.V.-----.-Y- -I-------.--H-V-N---..--..---....-R...-.---.KN-.-----.---T-M-.--A-.-------
B_GB4.M3015617  -----A-------I---------V-- ----------.R-SM..-----..VY-----..-.-.----K.--- -I--ED--K.----------..--..--V....--...-.---.---.-----.---T---.----.-------
B_GB4.M73767719 -----D--S----------K---Q-- ----------.----..---E..-LF--.D..-.-.-N---.--- ----TQ---.-----AI---.S-..--R....--...-.---.---.----L.-L---T-.----.------S
B_GB4.M73768520 --L------N---------K-A---- ----S-----.--P-..-----..-W-A----..-.-.----K.--- TIN------.------G---.G-..--V....--...-.---.-H-.-----.-------.----.------S
B_GB4.WB29      --------N-I-N---------- ----G----R.--T-..-----..--HA--R..-.-.----R.--- -----Q---.-----AT--..--..---E....--...-.---.---.-----.-V-L-I-.----.-------
B_GB5.4656      -------------------KVP-D-- ----X-----.----..-----..---A--G..-.-.----X.--- -I--V----.--I-X-X---..--..---....---...A.-X-.KP
B_GB5.4657    See inserts
B_GB5.4658      -----S-------V-------T---- ----G---SX.G---..-----..---A--G..-.X.----X.-X- ---S----D.---R-AX--..--..---N...--...-.-A-.--
B_GB5.4659      -----A-------I------Q----- ----------.----..-----..---A---..-.-.-----.--- --------D.--Q---I---.K-..--K....--...-.---.K-
B_GB5.4660      I--X----SN---------K-P-Q-- ----------.--T-..-----..T---.-D..-.-.-----.--- -----D--S.------A---.K-..K-E....D-...-.---.K-
B_GB5.4661      ---------N---------S-A---- --------X-.G---..-----..TW----Q..-.-.------.-Y- -I--V----.--GR--E---..--..---....---.-.---.-R
B_GB5.4662      -----V-------I-------T-Q-- ---------R.----..---S..-V----G..-.-.-N---.--- -I---Q---.--G-------..--..--K....--...-.---.-H
B_GB5.4663      I------I-N---S-L-H-------- ----------.----..-----..---A-.D..V.-.-----.-Y- -I-G-----.------T---..--..K--....--...-.-----.--
B_GB5.4664      --------------------A----- ----------.G---..-----..---A---..-.-.-----.--- ----KM--D.--QR--I---.G-..--X....--...-.-I-.XP
B_GB5.4665      -----D-LM--T-------K---N-T -V---------.--N-..-----..------D..-.-.-----.--- -I-KKG---.--Q-VA----.ST..--D....--...-.---.-X
B_GB5.4666      -----D---------L---TK--N-T -V---------.--N-..-----..------Q..-.-.-----.-Y- -I-KT---D.--Q--AR--..ST..---....--...-.---.--
B_GM.GM6          ------N---V---H-----K-D -I---------.--P-..-----..--F---D..-.-.---G-.--- ---IV---D.---R--T--..--..--K....--...-.---.-S-.----L.-V-T-
B_HT.RF         ----------V--------A--Q-- ----------.--TK..-----..VI-A--Q..-.-.----K.--- -----Q---.----V-T--..--..--D....--...-.---.TS-.-----.---L---.----.------
B_HT1.D1HA65    -I-------------------T-IQ-- ----------.--N-..-----..-W-A-.N..-.-.-----.--- ---G-E--K.A---L-T--..--..---....--...-.---.-S-.----L.-------.----.-------
B_HT1.D1HA652   I--------N----------IA-T -I---------.--P-..-----..---A--D..-.-.-----.--- ---I-N-TD.--NR-AT--..--..KY-H...--...-.---.KPH.A----.---T---.----.------
B_HT1.D2HA590    G---K-----TQ----H--Q--I-- ----------.----..-----..---.---..-.-.-----.--- ---K-S-T-.A-R---T---.G-..-LRH...-T...-.-A-.QPPI-----.-------.----.---
B_HT1.D2HA593   ---------N------I----T---- --------S-.R-S-..-----..--RA-.K..-.-.-N---.--- -I---T--S.----K--A--..--..---....--...-.---.QP-.-----.-------.----.-------
B_HT1.D2HA594   ---------N------I-PK---V-- ------S----.R-S-..-----..VW----Q..-.-.-----K.--- ---STQ--D.--G---E---..-A..---....--...-.-R-.--P.-E---.---L---.----.-------
B_HT1.D2HA596   ---------N------I--K---V-- ----------.R-S-..-----..VW----Q..-.-.-N--K.--- ---STA---.--G---E--..-A..---....--...-.-R-.--P.-----.---L---.----.-------
B_HT1.D2HA599 See inserts
B_HT2.H13954     G---K-----TQ----H--Q--I-- ----------.----..----..---.---..-.-.-----.--- ---K-S-T-.A-R---T---.G-..-LRH...-T...-.-A-.QPPI-----.-------.----.---
B_HT2.H13958     --------N------I----T---- --------S-.R-S-..-----..--RA-.K..-.-.-N---.--- -I---T--S.---K--A--..--..---....--...-.----.QP-.-----.-------.----.---
B_HT2.H13960     --------N------I--K---V-- ----------.R-S-..-----..-S----Q..-.-.----K.--- ---STG---.--G---E---..-A..---....--...-.-R-.--P.-----.---L---.----.----
B_HT2.H13962     --------N------I--K---V-- ----------.R-S-..-----..-W----Q..-.-.----K.--- ---STQ---.--G---E---..-A..---....--...-.-R-.--P.-----.---L---.----.----
B_HT2.H13968   See inserts
B_HT2.H5986      .K----------I-------A----- ----S-----.--PM..---K..------D..-.-.-----.-Y- -I-G-----.---R-----..--..---....-TT..-.-I-.-S-.-----.-----T-.----.------
B_HT2.H5988      ----------G--------N----- ----------.-----..-----..T-F---D..-.-.-----.--- -----Q---.---H--E---..E-..---....--...-.---.---.-----.-------.----.-------
B_HT2.H5990      --------N----------Q----- ----------.--N-..-----..---A---..-.-.-N--E.--- -I--T---E.--R---A---..-D..---....--...-.---.---.-----.-------.----.-----D
B_HT2.H5992      ----------T----------T -E------KR.--T-..-----..--DAY-G..-.-.-----.--- -V-I-E---.----------..--..--R....--...-.---.T--.-----.-------.----.------
B_HT2.H5994      -------S----------KDP-Q-- ----------.---M..---K..--------..-.V.-----.--- -----E---.A-T---T--..--..---....--T..-.-A-.KN-.-----.---T---.----.------
```

V3 LOOP
B_CONSENSUS_96 VVIRSENFTDNAKTIIVQLNESVEIN CTRPNNNTRK.SIHI..GPGR..AFYTTGE..I.I.GDIRQ.AHC NLSRAKWNN.TLKQIVKKL..RE..QFG....NK...T.IVF.NQS.SGGDP.EIVMHSF.NCGG.EFFYCNT
B_HT2.H5996
B_HT2.H5998
B_HT2.H6000
B_HT2.H6002
B_HT2.H6004
B_HT2.H6006
B_HT2.H6008
B_HT2.H6010
B_HT2.H6012
B_HT2.H6014
B_HT2.H6016
B_HT2.H6018
B_HT2.H6020
B_HT2.H6022
B_HT2.H6024
B_ID.1701
B_IN.IND9
B_IN1.20016 See inserts
B_IN1.20023 See inserts
B_IN1.30005 See inserts
B_IN1.30008
B_IT,PD See inserts
B_IT.Sala
B_IT.Sala2
B_IT1.115
B_IT1.130
B_IT1.136
B_IT1.145
B_IT1.190
B_IT1.193
B_IT1.196
B_IT1.199
B_IT1.20
B_IT1.204
B_IT2.13A
B_IT2.17B
B_IT2.32B
B_JP.ETR
B_JP.GUNA
B_JP.JH32
B_JP.JNIH1M
B_JP.KM03
B_KP.Kr111
B_LT.LIT11A
B_LT.LIT17A
B_LT.LIT18A
B_LT.LIT21A
B_LT.LIT22A
B_MY1.MRN02
B_MY1.MRNP05
B_MY2.9214087
B_MY2.9214089
B_MY2.9214093
B_MY2.9214096
B_MY2.9315157
B_MY2.9315158
B_MY2.9315160
B_MY2.9315168
B_MY2.9315171
B_MY2.9315172
B_MY2.9315174

```
                                                  V3 LOOP
                  ^^^          ^^^   ^ **   ^^^                    * ^^^    ^^ ^                    ^^   ^     ^^^                 *       *
B_CONSENSUS_96  VVIRSENFTDNAKTIIVQLNESVEIN CTRPNNNTRK.SIHI..GPGR..AFYTTGE..I.I.GDIRQ.AHC NLSRAKWNN.TLKQIVKKL..RE..QFG....NK...T.IVF.NQS.SGGDP.EIVMHSF.NCGG.EFFYCNT
B_MY3.1748      -----S------RV---------- ----------.---L..---Q..-W----Q..-.-.-----.--- ---ST----.-----TE--..-
B_MY3.1755      -----S------RV----------D ----------.--PL..---Q..-W----Q..-.-.-----.--- ---ST----.-----TE--..-
B_MY3.1763      -----S------RV---------- -I--------.---L..---Q..-W----Q..-.-.-----.--- ---ST----.-----TE--..-
B_MY3.1782      -----S---N--RV---------- ----------.--PL..---Q..-W----Q..-.-.-----.--- T--ST----.-----YE--..-
B_MY3.M1739     -----S------RV---------- ----------.---L..---Q..-W----Q..-.-.-----.--- ---ST----.--R--T---..-
B_NL.114C       --------------------KT---- ----------.--P-..----..-W----D..-.-.-----.--- ---ITQ--K.--E--AM--..--..--K....--...-.---.--
B_NL.168        I--------N---I------V--D-- -------I--.R---..----..------Q..-.-.-N---.--- -I-KTR---.--N--A---..--..K-....--...-.---.---.------.-------.----.-------
B_NL.Xlcon    See inserts
B_NL.h159       -----N----------------IA-- ----------.----..----..------D..-.-.-N---.--- ----T--E-.-----A---..--..---....--...-.---.-H
B_NL.wolfscon   -----D----------------V-- ----------.----..----..------D..-.-.-----.---
B_NL1.A         I----D-I----------K-A-Q-- ----------.----..---K..---A---..-.-.-----.--- ----VD-ED.-----AE--..--..--R....--...-.---.--
B_NL1.A10       -----A-------------------- ----------.----..---S..------D..-.-.-----.--- -I--T-----.--Q------..--..--N....--...-.---.--
B_NL1.A13       -I---A---N----------K----- ----------.RVTM..-----..VW-----..-.-.-N-K-.--- -----E-E-.-----VAE--..-G..---N....--...-.-A-.--
B_NL1.A2        ---------N-----M---K-P-V-- ----------.G--L..---G..T--A---..-.-.-----.--- ---ST---D.--R---T--..--..--V....--...-.---.-R
B_NL1.A3        -A---S-------------------- -I--------.--N-..----..-------..-.-.-----.--- -I-G----S.---M--E--..--..---....--...-.-N-.--
B_NL1.A4        I--------------------K----- ----------R.G---..-L---..R---R.K..-.-.-----.--- -I--------.--R---D--..--..---....--...-.-I-.--
B_NL1.A5        ---------N-------------V-- ----H-----.----..----..-------..-.-.-----.--- ---SV----.--R---R--..K-..---....--...-.---.--
B_NL1.A7        ---------N----------T---- ----G----R.--P-..---K..--F--.-..-.-.-----.--- -I-KG----.M---V-I--..--..---E....--...K.---.--
B_NL1.A9        -----A-------------------- ----------.-----..-L---..------D..-.-.-----.--- -I-GT----.--Q------..--..--N....--...-.---.--
B_NL10.BR8914   --V------N------I------A-S ----------.---L..-L---..---A---..V.-.----K.-Y- E-NGT----.-----AR--..K-..---....--...-.-E-
B_NL10.GF921953 ---------N----------T----- ----------.-VP-..----..-----.-..-.-.-N---.--- -----T--A.A-NKT-E--..--..---K....--...-.-T-
B_NL10.GQ93013  --L--D-------------K------ ----S-----.--P-..----..------D..-.-.-----.--- -I-GT---D.A----AQ--..--..---....--...-.---
B_NL10.MA9136   -----K-----T--------KP---- --------M-.----..----..------D..-.-.-----.--- -I---E--D.-----AG--..--..---V....--...-.---
B_NL10.SR9038   I----A-LSN---------K-P-Q-- ---------R.--P-..---K..V--.--D..-.-.-----.--- ----T---E.------G--..--..---....--...-.---
B_NL10.SR9115   -----K-------S---------V-- ---------R.--PM..-----..V--A--D..-.-.-----.--- -I--T-----.A-----I--..G-..---....--...-.-A-
B_NL10.SR9235   I----K-I----------K-PI--- ----------.--PM..----..------Q..-.-.-----.--- ---IE---D.-----AD--..--..KV-N...--...-.-I-
B_NL10.SR925752 ---------N----------V----- ----------.--N-..----..------Q..-.-.-----.--- -I--E----.----V-T--..K-..---K....--...-.-I-
B_NL10.SR9269   -----D--SN---N-------T---- ----------.----..----..-------..-.-.----R.-Y- -IN-T-----.--E--AR--..K-..--.....--T..-.V--
B_NL10.SR93047  -----V----TG-N---------V-- ---------R.---F..---A..R--A---..-.-.-----.--- -I-G-Q---.---W--MR-..--..---....--...-.---
B_NL10.UM893272 --------------------T---- ---------R.--P-..----..-W----D..-.-.-----.--- -----Q--K.--E--AM--..--..--K....--...-.---
B_NL10.UM94030  ----------K--V-------- ----------.----..----..---A--D..-.-.-----.--- --------E.----V-T--..-A..--R....-R....-.-S-
B_NL10.UM94038  ----------VR--------T---- ----------.----..----..---A--D..-.-.-N---.--- -------EK.--E---I--..G-..---....--...-.---
B_NL11.patFcon  -------L----------------V-- ----S-----.--PL..----..---A-.D..-.-.-----.--- -I--TE---.------T--..--..--K....--...-.---.-H-.------.----
B_NL11.patNcon  -----A---N---V---------- ----G-----.R-SL..-----..VY----Q..-.-.-----.--- -I--------.--R--AG--..K-..---....--...E.---.TH-.------.--A
B_NL12.0008     -----D----------K------ ----------.--X-..----..-------..-.-.-----.--- ----T--D-.-----AR--..--..---....--...-.---.--
B_NL12.0157     -----D-------I-----K-P---- ----------.--N-..----..-------..-.-.-----.--- -I-ST---D.X-R---R--..--..---....--...-.-G-.-R
B_NL12.0211     ---------N--R------------ -I--------.----..----..-------..-.-.-----.--- ---ST---D.---------..--..---....--...-.---.--
B_NL12.0230     --L--D-----------A------ ----------.----..----..------G..-.-.-----.--- ---K-----.A-E--A---..-G..--E....--...-.---.--
B_NL12.0337     -----D----------K-----X -X--------.----..----..-------..-.-.-----.--- TI---N-T-.A-Q------..--..E-K....--...-.---.--
B_NL12.0583     ---------N---------K---- ---------X.--N-..----..-------..-.-.-----.--- ---SL---D.--R---V--..--..---....--...-.---.-K
B_NL12.0617     ---------N---------K---- ----------.--N-..----..-------..-.-.-----.--- ----T--D-..-E------..--..--K....--...-.---.--
B_NL2.Wolfsh     ----D-I----------KKA-Q-I ----------.--P-..----..---A---..-.-.-----.--- -----VN-EI.--H--AE--..--..--R....--...-.-A-.--
B_NL2.Wolfsh495  ----D-------I------------ ----------.--P-..----..-------..-.-.-----.--- -I--------.--G------..--..--R....--...-.-I-.--
B_NL3.NET       ----------T--------K----- ----------.----..----..-------..-.-.-N---.--- -I--------.-----AV--..--..--K....--...-.---.-H
B_NL3.NET2      ---------N---------K-P-R-- ---------R.--G-..----..-I-A-DR..-.V.-N---.--- ---K-E--R.--E------..--..--R....--...-.---.--
B_NL3.NET3      -----D--S----------KDP-N-T -I--------.--S-..----..---A---..-.-.-N---.--- ---GTE---.A---VAE--..SK..---....--...-.---.K-
B_NL3.NET4      -------L-----------V----- ----------.G---..----..---A-ER..-.-.-N---.--- -I-SE----.---LV-T--..--..K--N...--...-.---.--
B_NL3.NET5      --------SN---------K-P-K-- ---------R.G---..----..-V----K..-.-.-----.--- -I--T---D.--Q---I--..K-..---....--...-.---.--
B_NL3.NET6      --M-----N---S------------ ----S----R.----..----..------Q..-.T.-N---.--- S---------.------I--..-K..--E....-R....-.---.--
B_NL4.B63_84     ----D-------I-----K-P---- ----------.--Q-..----..-------..-.-.-N---.--- -I--------.--Q---I--..--..---....--...-.---.---
B_NL4.BA1_84     ----A---N---------K-P---- ----------.----..----..------D..-.-.-----.--- ---SS--YD.--R-VA---..--..---....--...-.---.---
B_NL4.BA206_84   ----D-------------X---V-- ----------.----..----..-------..-.-.-----.--- --------D.--E-V-X--..--..---    S-...-.---.---
B_NL4.BA65_85    ----D-------------KAP-V-- ----------.G---..----..-------..-.-.-----.--- ----TQ--D.P-Q--AI--..--..--K....--...-.---.---
B_NL4.BN130_85   ----D---------------------- ----------.----..----..---A---..-.-.-----.--- ---K---E-.--E------..--..--K....-ET..-.---.---
B_NL4.BN132_84   ----V--A----------G------ ----------.RMTL..---K..V-------..-.-.-----.--- -------D-.-----AT--..--..--V....--...-.---.-R-
B_NL4.BN142_84   --------N---I------------ -I--------.----..----..-------..-.-.-----.--- -I-K-R---.--E------..--..---....--...-.---.---
B_NL4.BN152_84   ----D---N--N------K------ ----------.----..----..-------..-.-.-----.--- ----E--D-.A-E--DG--..--..--S....--...-.---.---
B_NL4.BN18_84    ----D--S----------A----- ----------.----..----..---A----..V.-.-----.--- ----T--D-.I-E--A-N-..--..--R....--...-.---.-H-
B_NL4.BN29_85    ----D------R-------A--V-- ----------.G---..----..------D..-.-.-----.--- ----TE-D-.-----AT--..--..---....--...-.-I-.---
B_NL4.BN90_85    I------------------T----- ----------.----..----..-------..-.-.-----.--- -I--E-----.M-Q---I--..--..--E....-R....-.---.KR-
B_NL4.BN93_85    --------N---I------------ -I--------.----..----..-------..-.-.-----.--- -I-K-----.---------..--..--V....--...-.---.---
```

V3 LOOP
B_CONSENSUS_96 VVIRSENFTDNAKTIIVQLNESVEIN CTRPNNNTRK.SIHI..GPGR..AFYTTGE..I.I.GDIRQ.AHC NLSRAKWNN.TLKQIVKKL..RE..QFG....NK...T.IVF.NQS.SGGDP.EIVMHSF.NCGG.EFFYCNT
B_NL4.H1091_90
B_NL4.H1112_90
B_NL4.H135_90
B_NL4.H153_82
B_NL4.H164_90
B_NL4.H169_85
B_NL4.H17_82
B_NL4.H186_85
B_NL4.H187_90
B_NL4.H188_90
B_NL4.H1_85
B_NL4.H228_90
B_NL4.H23_82
B_NL4.H293_85
B_NL4.H29_82
B_NL4.H339_82
B_NL4.H35_80
B_NL4.H36_82
B_NL4.H434_85
B_NL4.H43_90
B_NL4.H450_85
B_NL4.H457_85
B_NL4.H466_82
B_NL4.H490_85
B_NL4.H495_85
B_NL4.H552_90
B_NL4.H557_82
B_NL4.H594_85
B_NL4.H724_8
B_NL4.H748_90
B_NL4.I1020_86
B_NL4.I1025_89
B_NL4.I1035_87
B_NL4.I1057_87
B_NL4.I1083_87
B_NL4.I1164_90
B_NL4.I117_90
B_NL4.I138_89
B_NL4.I141_9
B_NL4.I142_87
B_NL4.I221_90
B_NL4.I257_9
B_NL4.I3007_86
B_NL4.I3008_88
B_NL4.I3048_89
B_NL4.I3057_9
B_NL4.I3064_90
B_NL4.I3072_90
B_NL4.I3077_9
B_NL4.I3084_90
B_NL4.I3091_9
B_NL4.I33_90
B_NL4.I4004_88
B_NL4.I4008_86
B_NL4.I5020_89
B_NL4.I5032_87
B_NL4.I5037_89

V3 LOOP
B_CONSENSUS_96 VVIRSENFTDNAKTIIVQLNESVEIN CTRPNNNTRK.SIHI..GPGR..AFYTTGE..I.I.GDIRQ.AHC NLSRAKWNN.TLKQIVKKL..RE..QFG....NK...T.IVF.NQS.SGGDP.EIVMHSF.NCGG.EFFYCNT
B_NL4.I5038_89
B_NL4.I7025_9
B_NL4.I70_88
B_NL4.I90_89
B_NL4.K999_86
B_NL5.pt10con
B_NL5.pt11con
See inserts
B_NL5.pt12con
B_NL5.pt13con
B_NL5.pt14con
B_NL5.pt15con
B_NL5.pt16con
B_NL5.pt17con
B_NL5.pt19con
B_NL5.pt1con
B_NL5.pt2con
B_NL5.pt3con
B_NL5.pt4con
B_NL5.pt5con
B_NL5.pt6con
B_NL5.pt7con
B_NL5.pt8con
B_NL5.pt9con
B_NL6.16
B_NL6.320
B_NL7.NE1058con
B_NL7.NE537con
B_NL8.594
B_NL8.672
B_NL9.H239
B_NL9.h1054
B_NL9.h1081
B_NL9.h1082
B_NL9.h1110
B_NL9.h1135
B_NL9.h1140
B_NL9.h1145
B_NL9.h159
B_NL9.h39
B_NL9.h411
B_NL9.h412
B_NL9.h424
B_NL9.h526
B_NL9.h569
B_NL9.h57
B_NL9.h571
B_NL9.h709
B_NL9.h716
B_NO1.10
B_NO1.11
B_NO1.12
B_NO1.15
B_NO1.16
B_NO1.20
B_NO1.21
B_NO1.22
B_NO1.23
B_NO1.24
B_NO1.25

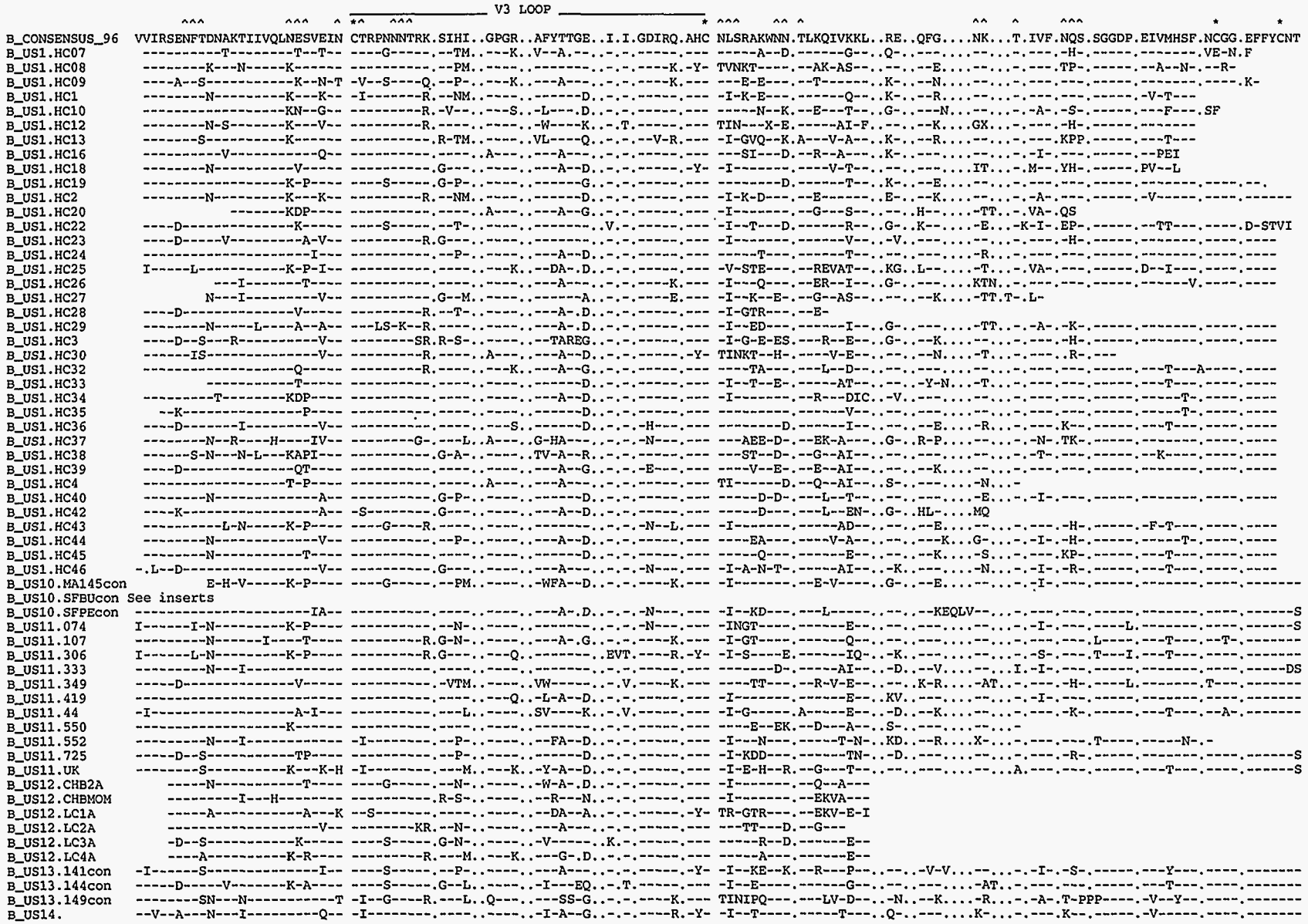
V3 LOOP
B_CONSENSUS_96 VVIRSENFTDNAKTIIVQLNESVEIN CTRPNNNTRK.SIHI..GPGR..AFYTTGE..I.I.GDIRQ.AHC NLSRAKWNN.TLKQIVKKL..RE..QFG....NK...T.IVF.NQS.SGGDP.EIVMHSF.NCGG.EFFYCNT
B_NO1.26
B_NO1.27
B_NO1.29
B_NO1.30
B_NO1.31
B_NO1.32
B_NO1.34
B_NO1.35
B_NO1.36
B_NO1.37
B_NO1.38
B_NO1.39
B_NO1.40
B_NO1.41
B_NO1.42
B_NO1.43
B_NO1.44
B_NO1.45
B_NO1.O1
B_NO1.O3
B_NO1.O4
B_NO1.O5
B_NO1.O6
B_NO1.O8
B_NO1.O9
B_PR1.D2PR728
B_PR1.D2PR729
B_PR1.D2PR732
B_PR1.D2PR733
B_PY.12837
B_PY.12838
B_PY.12839
B_PY.1284
B_PY.12842
B_PY.12845
B_PY.12847
B_PY.3614p
B_PY.3615
B_PY.3616
B_RU.RUS3A
B_RU.RUS4A
B_RU2.10120
B_RU2.154
B_RU2.190
B_RU2.20637
B_RU2.2236
B_RU2.3818
B_RU2.4035
B_RU2.4112
B_RU2.4261SE
B_RU2.4439
B_RU2.4466
B_RU2.4714
B_RU2.5014
B_RU2.5852
B_RU2.999
B_SE.pt11s113
B_SE1.1032
B_SE1.1433
B_SE1.1866
B_SE1.2815

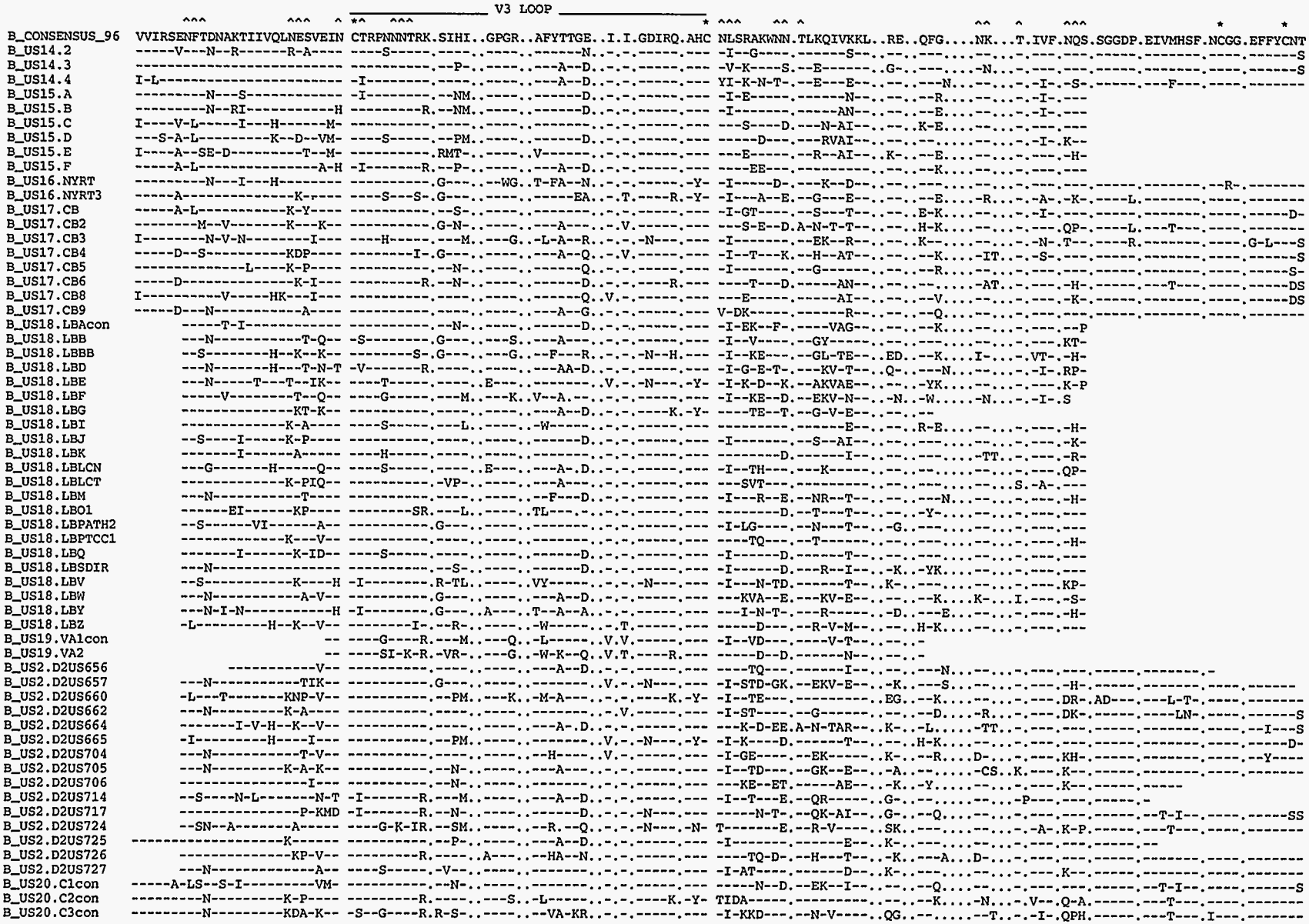
V3 LOOP
B_CONSENSUS_96 VVIRSENFTDNAKTIIVQLNESVEIN CTRPNNNTRK.SIHI..GPGR..AFYTTGE..I.I.GDIRQ.AHC NLSRAKWNN.TLKQIVKKL..RE..QFG....NK...T.IVF.NQS.SGGDP.EIVMHSF.NCGG.EFFYCNT
B_SE1.295 ----D----------KDPIK-- ---------R.--N-..----..---A---..-.-.------.--- ---E-A--K.-----AEG
B_SE1.93 -------MN---------K-P---- ----------.----..----..---A--A..-.-.-N---.--- IIDKV-----.-----Y---
B_SE1.930 ------I---T--------KT-T-- ----------.----..A---..---A---..-.-.------.-Y- -I--TE---.--G-----
B_SE2.R1 ----- ----------.----..----..-------D..-.-.-N---.--- -------E-.-------E-..--..--K....--...-.---.-
B_SE2.R2 P---- ----------.--S-..----..-------D..-.-.------.-Y- ---I------.------G--..--..--K....--...-.---.-
B_SE2.R3 -I--- ----------.--P-..----..-------D..-.-.------.--- ---K-T---.--E-M-T--..--..--W....-N...-.-AL.R
B_SE2.R4 P---- ----------.--P-..----..---A--D..-.-.-Q---.--- IIDK-----.----------..--..--N....--...-.---.-
B_SE2.R5 ----I -V---------.--T-..----..--------..-.-.-N---.-Y- -----A--E.--R-VA----..K-..--.....-R...-.---.--
B_SK1.BTS11 -----D-L---T-------KD--K-- ----------.G---..---K...--A---..-.-.-N---.--- -I---E-EK.A-N-TSI--..S-..H-K....--...-.-A-.---
B_SK1.BTS12 ---------N----------N-IA-E ----S-----.G--M..---K...-FA---..-.-.------.--- -I-ETD-E-.A---VAI--..S-..H-P....--...-.-I-.---
B_SK1.BTS15 See inserts
B_SK1.BTS18 -----D----------K------ ----------.----..----..--------..-.-.-N---.--- TI-KT--YS.--N-T-I--..--..---....--...-.---.---
B_SK1.BTS20 -----A----------S---K-- ---------R.---L..---K..---A-.D..-.-.------.--- -----D--K.--N---T--..--..---N...-R...-.-I-.---
B_SK1.BTS22 -----A----------E---Y--T-- ----------.--R-..---S..T--A---..-.T.------.--- ----EE---.--N-------..K-..----...--...-.---.---
B_SK1.BTS23 --------A---------------- -I-------R.-----..-----..--------..-.-.--V-R.--- -I--E----.-----AI---..S-..R-N....--...-.---.---
B_SK1.BTS28 ---------N---------K-P---- ----G----R.--N-..----..------A..-.-.------.--- -V-KE----.--N------..--..---....-R...-.-I-.-H-
B_SK1.BTS9 ----------K-P-V-- ----------.-V-L..--K-..V-F---K..-.V.------.--- ----TQ----.--E------..--..---....--...-.-S-.---
B_TH.93TH067 -----S------RV---------- ----------.--PL..---Q..-W----Q..-.-.------.--- ----GT--R-.--E--TE--..K-..---....--...-.---.---.------.---R-T-.-----.-------
B_TH.JP23B ---N----------D----- -S--------.---V..----..-L-----..-.-.------.--- -----Q--D.--HK--IT--..--..---....--...-.---.KH-.-
B_TH.T8174 -S-----TRI---------- ----------.--PL..---Q..-W----Q..-.-.------.--- ---ST----.-----TE--..--..---....--...-.---.---.------.-------.-----.-------
B_TH1.7787 ---TRI---------- ----------.---L..-----..-W----Q..-.-.------.--- ---ST----.-----TE--..--..---....--...-.---.---.---
B_TH1.864 -V--S-------V---------- ----------.---L..---Q..-W-A---..-.-.------.--- ---ST----.--R--T---..--..---....--...-.---.---.------.-------.-----.------
B_TH1.8643 ----S------RV---------- ----------.---L..-Q-Q..-W-A---..-.-.------.--- ---ST--I-.------TE---..-D..---....--...-.---.---.------.-------.-----.------
B_TH1.8645 ----S-------V---------- ----------.---L..---Q..-W----Q..-.-.------.--- ---ST----.-----T---..--..---....--...-.---.--P.----Q.-------.-----.------
B_TH1.8647 ----S------RV---------- ----------.--PL..-----..-W-A---..-.-.------.--- ---ST----.--R--TE--..--..---....--...-.---.---.------.-------.-----.------
B_TH1.8649 -----RV---------- ----------.--PL..----Q..-W----Q..-.-.------.--- ---ST----.-----AE--..--..---....--...-.---.---.------.-------.-----.------
B_TH1.865 ------RV---------- ----------.G---L..---Q..-W----Q..-.-.------.--- ---ST--H-.--E--TE--..--..---....--...-.---.---.------.-------.-----.------
B_TH1.8653 ----S------RV---------- ----------.--PL..---Q..-W----Q..-.-.------.--- ---ST----.-----TE--..--..---....-T...-.---.---.------.-------.-----.------
B_TH1.8655 -S----N--RV---------- ----------.--PL..---Q..VWF---Q..-.-.------.--- ---ST----.-----TE--..--..---....--...-.---.--P.-----.-------.-----.------
B_TH1.8669 -----RI---------- ----------.---L..---Q..-W----Q..-.-.------.--- ---ST----.-----T---..--..---....--...-.---.-A---.-------.-----.------
B_TH2.CM237 ---K-S---G--RV-----D------ ----------.---L..---K..-W----P..-.-.------.--- ---ST--T-.-----TE--..--..---....--...-.---.---.------.-------.-----.-------
B_TH2.TB132 -X----------K-P---- ------Y----.X-XM..-----..XX-----..-.-.----R.--- -I-ST----.--G-------..K-...-N....-N...-.---.-E-.-----.-------.I-W-.---
B_TH3.W2TH014 ------RV---------- ----------.---L..-----..-W----Q..-.-.------.--- ---ST----.--R--TE--..--..---....--...-.---.---.-
B_TH3.W2TH026 ---N---I---------- ----------.--PL..---Q..-W----Q..-.-.------.--- ---ST----.--R--TE--..--..---....--...-.---.---.-
B_TH4.N752 ---T----------I--- ----------.G-P-..----..------D..-.-.------.--- ----TQ-D-.A----A----..G-..-L-....-TT..-.---.---.------.---A---.-----.------S
B_TH4.N753 -----I------D-IQ-- --------S-.G-P-..----..---A--X..-.-.------.--- -I--TD-KK.A-----E---..G-..-YR....-N...-.-A-.---.----L.-V------.-----.-------
B_TH4.N754 ---T----------I--- ----------.G---..----..--------..-.-.------.--- -I--T--D-.A----AEQ-..--..-L-....--...-.---.---.----L.---T---.-----.-------
B_TH4.N755 ---T----------IV-- ----------.G---..----..--------..-.-.------.--- -I--T--D-.A-----EQ-..--..-L-....--...-.---.---.----L.---T---.-----.-------
B_TH4.N756 ---T---------A-I--- ----------.G---..----..--------..-.-.------.--- ----T--D-.A----A---..--..-L-....-N...-.---.---.------.---T---.-----.------S
B_TH4.N757 ---T---------A-I--- ----------.G---..----..--------..-.-.------.--- ----T--D-.A---------..--..-L-....--...-.---.---.------.---T---.-----.------S
B_TH4.N758 ---T---------Q-I--- ----------.--P-..----..-------Q..-.-.------.--- -------ED.A----AT--..--..-L-....-Q...-.---.---.------.---L---.-----.------P
B_TH4.N759 ---T---------K-I--- ----------.G---..----..--------..-.V.------.--- ----TE-D-.------G---..--..-L-N....--...-.---.---.------.-------.-----.------S
B_TH4.N76 -N-V-N---------- ----------.---M..-Q---..---A--G..-.-.------.-Y- TI-GE----.--Q---T--..--..--K....-R...-.-A-.---.T----.-------.-----.-------
B_TH4.N760 ---T-------KXP---- ----------.G---..----..------G..-.-.------.--- -------Q---.--D-------..--..H--....--...-.---.---.----L.-V-T---.-----.-------
B_TH4.N762 ----RV---------- ----------.G--L..---Q..-W----Q..-.-.------.--- ---ST----.-----AE--..--..---....-N...-.---.---.------.-------.-----.-------
B_TH4.N763 ----RV------D----- ----------.G-PL..---Q..-W----Q..-.-.------.-Q- ---ST--T-.--R--TE--..--..---....--...-.---.---.------.-------.-----.-------
B_TH5.018 --L--A---E-T-I---------V-S ----------.----..----..------D..-.-.------.--- -I--KD--S.--RR-AI--..--..---....--...-.---.KP-.A----.-------.--K-.-------
B_TH5.1948 -A----------------K-P---- ------Y----.R-TM..-----..VY------..-.-.----R.--- -I-ST----.--G-------..K-...-N....-N...-.---.KK-.------.-------.I---.---F--S
B_TH5.2619 -I---S------RI---------V-- ----------.--PL..---Q..-W----Q..-.-.------.--- ---ST----.-----AE--..--..---....--...-.---.---.------.-------.-----.---H---
B_TH6.0258 -----S-----TRI---------- ----------.---L..---K..-W----Q..-.-.------.--- ---ST----.A-R--TE--..-
B_TH6.0286 -I---A--S----I-----KDP---- ----------.--P-..---...--FR---..-.-.----K.--- -I-KT----.--Q-V-G--..-
B_TH6.1611 -----K-I---T-I-----KDP---- ----------.G---..---K..--------..-.-.------.--- ---K-S---.----------..-
B_TH6.1615 ----------T-N-----K--I--- ----------.----..-----..------Q..-.-.------.-Y- ---K-N-EE.A-ER-AT--..-
B_TH7.101 --S----------A---- ----------.RVTL..---K..VW----Q..-.-.------.--- ---ST----.--R--AE--..--..---....--...-.---.---.----L.-------.-----.-----D-
B_TH7.102 -------I----------K ----------.---L..-Q---..-W----R..-.-.------.--- ---STD-SS.--N--A----..G-..--V....--...-.-A-.---.-----.-------.-----.-----D-
B_TH7.103 ---N--RV---------- ----------.G--L..---Q..-W----Q..-.-.------.--- ---ST----.---T-------..Q-..----....--...A.---.---.-----.-------.-----.-------
B_TH7.104 ------RV---------- ----------.--PL..---Q..-W----Q..-.-.------.--- ---ST----.-----TE--..--..---N...-T...-.V--.-H-.---P-.V--I---.-----.-------
B_TH7.105 ------RV---------- ----------.--PL..---Q..-WF---Q..-.-.------.--- ---ST----.-----A---..--..---....-T...-.V--.-H-.---H-.V------.-----.-------
B_TH7.108A -------V---------- -I--------.GTQ-..-----..TW----Q..-.-.------.--- ---ST----.-----T-N-..--..---....--...-.---.K--.-----.--A----.-----.-------
B_TH7.109 ---N--RV-------A---- ----------.--PL..-L---..-W----Q..-.-.------.--- ---ST----.A----TE--..--..---....--...-.---.---.------.-------.-----.-------
B_TH7.110 ------RV---------- ----------.---L..-----..-W----Q..-.-.------.--- ---ST----.-----TE--..--..---....--...-.---.---.------.-------.-----.-------

V3 LOOP

```
                    ^^^         ^^^   ^ *^   ^^^                  * ^^^    ^^ ^                ^^   ^     ^^^                 *       *
B_CONSENSUS_96  VVIRSENFTDNAKTIIVQLNESVEIN CTRPNNNTRK.SIHI..GPGR..AFYTTGE..I.I.GDIRQ.AHC NLSRAKWNN.TLKQIVKKL..RE..QFG....NK...T.IVF.NQS.SGGDP.EIVMHSF.NCGG.EFFYCNT
B_TH7.112       ------RV---------- ----------.--QL..-Q--..-W----Q..-.-.-----.--- ---ST--S-.-----AE--..--..---....--....-.---.---.-----.------V.-----.-------
B_TH7.113       ------RV---------- ----------.--PL..---Q..-W----Q..-.-.-----.--- ---ST--S-.-----AE--..--..---....--....-.---.---.-----.-------.-----.-------
B_TH7.114       -------V---------- ----------.G--L..---Q..-LCA--Q..-.-.-----.--- ---ST----.--RR-TE--..--..---....--....-.---.---.----L.-------.-----.-------
B_TH7.116       -------V---------- ----------.--PL..---Q..-WF---Q..-.-.-----.--- ---ST----.-----AE--..S-..---    S-....-.A-..-S-.-----.-L-----.-----.-------
B_TH7.117       -----TRI---------- ----------.---L..---Q..-W----Q..-.-.-----.--- ---ST----.--R--TE--..--..---....--....-.---.---.----L.-V-----.T---.-------
B_TH7.118       -------V------K--- ----------.G-PL..----..-W----Q..-.-.-----.--- ---STE---.-----A---..--..---....--....-.---.---.----H.-------.-----.-----D-
B_TH7.120       ---N--RI---------- ----------.---L..----..-WW---Q..-.-.-----.--- ---ST----.-----A---..--..---    S-....-.---.T--.-----.-VA----.-----.-------
B_TH7.122       ------RV---------- ----------.--PL..---Q..-W----Q..-.-.-----.--- ---ST----.-----AE--..--..---....--....-.---.---.-----.-------.-----.-------
B_TH7.123       ------RV---------- ----------.---L..---Q..-W----Q..-.-.-----.--- ---ST----.--N--TE--..--..---....--....-.---.---.-----.-------.-----.-------
B_TH7.124       ------RV---------- ----------.G-P-..---Q..-W----Q..-.-.-----.--- ---ST----.-----TE--..--..---....--....-.---.---.-----.-------.-----.-----D-
B_TH7.125       ---N--RI---------- ----------.--N-..----..-W----Q..-.-.-----.--- ---ST--T-.-----TE--..--..---....--....-.---.---.----R.--I-----.-----.-------
B_TH7.128       ------RV---------- ----------.---L..---Q..-W----Q..-.-.-----.--- ---ST----.--R--TE--..--..---....--....-.---.---.-----.-------.-----.-------
B_TH7.129       -------V---------- ----------.G--L..-L--..-W----Q..-.-.-----.--- ---ST--H-.---K-TE--..--..---....--....-.---.---.-----.-------.-----.-------
B_TH7.130       -----TRV---------- ----------.---L..---Q..-W----Q..-.-.-----.--- ---ST----.-----TE--..--..---....-T....V--.---.---.-----.V------.-----.-------
B_TH7.131       ------RV---------- ----------.--SL..----..-W----Q..-.-.-----.--- ---ST--S-.-----TE--..--..---....--....-.---.---.-----.-------.-----.-------
B_TH7.132       -----T-V---------- ----------.---L..---Q..-W----R..-.-.-----.--- ---ST----.----VTE--..--..---....--....-.---.---.-----.-------.-----.-----D-
B_TH7.133       ------RV---------- ----------.---L..---Q..-W----Q..-.-.-----.--- ---ST----.--N--AE--..--..---....D-....-.---.---.-----.-------.-----.-----D-
B_TH7.134       -------V---------- ----------.-----..---Q..-W-A--Q..-.-.-----.--- ---ST----.-----A---..-A..---....--....-.---.---.-----.--AT---.-----.-----D-
B_TH7.137       ------RV---------- ----------.--PL..---Q..-W----Q..-.-.-----.--- ---ST----.-----AE--..--..--R....--....-.---.---.-----.-------.-----.-------
B_TH7.138       ---N--RV---------- ----------.--N-..---Q..-W----Q..-.-.-----.--- ---ST----.-----AE--..--..---....-E....-.---.K--.-----.-------.-----.-------
B_TH7.139       ------RV---------- ----------.--PL..----..-W----Q..-.-.-----.--- ---ST----.-----AE--..--..--V....--....-.---.K--.-----.-------.-----.-------
B_TH7.140       ---N---V-M-------- ----------.R---..---Q..-L----Q..-.-.-----.--- I-NST----.--R--TE--..--..---    S-....-.---.K--.-----.-VA----.-----.-------
B_TH7.141       ------RV---------K-D -I--------.--QL..----..-W-A--Q..-.-.-----.--- -I-STE---.-----IV--..--..---....--....-.---.-R-.-----.--K----.-----.-------
B_TH7.149       ---N--RV---------- ----------.--PL..---K..TW----Q..-.-.-----.--- ---ST----.A----AE--..--..---....--....-.---.---.-----.-------.-----.-------
B_TH7.150       ------RV---------- ----------.--Q-..----..-W-A--Q..-.-.-----.--- ---ST----.-----TE--..--..---....--....-.---.-S-.-----.--A----.-----.-------
B_TH7.151       ------RV---------- ----------.---L..---Q..-W----Q..L.-.----H.--- ---STE-S-.---L-TE--..--..---....--....-.---.---.-----.-------.-----.-------
B_TH7.152       ------RV---------- ----------.--PL..----..-W------..-.-.-----.--- ---ST----.---L-AE--..--..---....--....-.---.-R-.-----.-------.-----.-------
B_TH7.153       ------RV-----S------ ----------.--PL..---Q..-W----Q..-.-.-----.--- ---ST----.--R--AE--..--..---....--....-.---.---.-----.-------.-----.-------
B_TH7.154       -------I---------- ----------.G---..---Q..-W----Q..-.-.-----.--- ---ST----.-----TE--..--..---....--....-.---.---.----L.-------.-----.-------
B_TH7.155       ------RV-------T---- --------I-.G--L..---Q..-W----Q..-.-.-----.--- ---ST----.-----AA--..--..---....--....-.---.---.-----.-------.T---.-------
B_TH7.156       ------RV---------- ----------.---L..---Q..-W-A--Q..-.-.-----.--- ---ST----.-----TG--..--..---....--....-.---.---.----Q.-V-----.-----.-------
B_TH7.157       ---N---V---------- ----------.--QL..----..-W-A--Q..-.-.-----.--- ---ST--H-.-----TE--..--..---....--....-.---.---.-----.--E--I-.-----.-------
B_TH7.158       -----TR----H------- ----------.-VPL..----..-W----Q..-.-.-----.--- ---ST----.--G--TE--..--..---....--....-.---.--P.----L.--Q--I-.-----.-------
B_TH7.159       ---N--RV---------- ----------.G--L..----..-W----Q..-.-.-----.--- ---ST--K-.-----TE--..--..---....--....-.---.---.-----.-V-----.S---.-------
B_TH7.160       ------RV---------- ----------.---L..---Q..-W----Q..-.-.-----.--- ---ST----.-----TE--..--..---....--....-.---.---.-----.-L-----.-----.------S
B_TH7.162       ---N--RV---------- ----------.---L..----..-W----Q..-.-.-----.--- ---ST----.--R--TG--..--..---....--....-.---.-H-.-----.-------.-----.-------
B_TH7.163       ------RV---------- ----------.-----..---Q..-W----Q..-.-.-----.--- ---ST--T-.-----T---..--..---....--....-.---.---.-----.-------.-----.-------
B_TH7.165       ------RV---------- ---------R.---L..----..-W-AP-D..-.-.-----.--- ---ST----.----VAE--..--..---....--....-.---.---.-----.-------.-----.-----D-
B_TH7.166       --S--TRV------K----- ----------.--PL..---Q..-WF---Q..-.-.-----.--- ---ITQ---.--E--TE--..--..R--....--....-.---.---.-----.-------.-----.-------
B_TH7.167       -------V---------- ----------.---L..---Q..-WF---Q..-.-.-----.--- ---STQ---.-----TE--..--..---....--....-.---.---.-----.-------.--A----.-----.-------
B_TH7.168       --S---RV---------- ----------.-V-L..----..-W----Q..-.-.-----.--- ---ST----.--N--A---..--..---....--....-.---.---.-----.-------.-----.-------
B_TH7.169       --S---RV---------- ----------.--PL..----..-W----Q..-.-.-----.--- ---ST--S-.-----TE--..--..---....--....-.---.---.----Q.-------.-----.-------
B_TH7.173       -----TRI---------- ----------.--PL..----..-W-A--Q..-.T.-----.--- ---ST--S-.-----AA--..--..---....--....-.---.---.----L.---L---.-----.-------
B_TH7.174       -----T-V---------- ----------.--PL..---K..-W----Q..-.-.----R.--- ---ST----.-----TE--..--..---....--....-.---.---.-----.-------T-.-----.-----D-
B_TH7.176       ---N---V---------- ----------.--PL..-A---..-W----Q..-.-.-----.--- ---ST----.M----TE--..--..---....--....-.---.---.-----.-------.---R.GI-----
B_TH7.177       ------RV---------- ----------.---L..---Q..-W----Q..-.-.-----.--- ---ST----.--R--TE--..--..---....--....-.---.---.----L.-------.-----.-------
B_TH7.181       ---N--RV---------A-- ----------.---L..---Q..-W----Q..-.-.-----.--- ---STQ---.---R-TE--..--..---....--....-.---.---.-----.-------.-----.-------
B_TH7.182       ------RV---------- ----------.--PL..---Q..-WF---Q..-.-.-----.--- ---ST----.--R--TE--..--..---....--....-.---.---.-----.-------.-----.-------
B_TH7.183       ------RI---------- ----------.---L..-Q---..-W-A--Q..-.-.-----.--- ---ST--H-.-----TE--..--..---....--....-.---.---.-----.-------.-----.-------
B_TH7.184       ---N--RV---------- ----------.--N-..----..-W-A--Q..-.-.-----.--- ---ST----.-----TE--..--..---....--....-.---.-R-.----L.--E----.-----.-------
B_TH7.82        --M---RI---------- ----------.--PL..---Q..-W----Q..-.-.-----.--- ---ST----.-----TE--..--..---....--....-.---.---.-----.-----N-.-----.-------
B_TH7.85        ------RV---------- ----------.--PL..---Q..-W------..-.-.-----.--- ---ST----.-----TE--..--..---....--....-.---.-H-.-----.-------.-----.-------
B_TH7.86        ------RI---------- ----------.--PL..---Q..-W----Q..-.-.-----.--- ---ST----.-----A---..--..---....--....-.---.---.----Q.-------.-----.-------
B_TH7.87        ---N--RV---------- ----------.--PL..---Q..-W----Q..-.-.-----.--- ---ST----.-----TE--..--..---....--....-.---.K--.----Q.-VE----.-----.-------
B_TH7.88        -------V---------- ----------.--PL..---Q..-W----Q..-.-.-----.--- ---ST----.A----TE--..--..--K....--....-.---.---.-----.-------.-----.-------
B_TH7.90        ---N--RV---------- ----------.--PL..---Q..-W----Q..-.-.-----.--- ---ST----.-----TE--..--..---....--....-.---.---.-----.-------.-----.-------
B_TH7.91        ---N---------K------ ----------.-----.A---..---A---..-.-.-----.--- -I-AE----.--G-V-A---.K-..-Y-S...-T...-.V--.D-P.AE--L.V-------.T---.-------
B_TH7.92        ------RV---------- ----------.--PL..---Q..-WF---Q..-.-.-----.--- ---ST----.-----AE--..--..---S...-T...-.V--.---.-----.---L---.-----.------S
B_TH7.93        --S----V--------T---- ----------.--PL..-----..-W----Q..-.-.-----.--- ---ST----.--R--TA---..--..---....--....-.---.---.-----.-------.-----.-------
B_TH7.94        -------V---------- ----------.--PL..---Q..-W----Q..-.-.-----.--- ---ST----.--R--A---..--..---....--....-.---.---.-----.-----T-.-----.-------
B_TH7.96        -----T-------T---K-D -I--------.--NL..-L---..-W-A--Q..-.-.-----.--- ---STE---.---K-TE--..--..-Y-....-TK..-.---.---.-----.-------.V---.-------
```

V3 LOOP
B_CONSENSUS_96 VVIRSENFTDNAKTIIVQLNESVEIN CTRPNNNTRK.SIHI..GPGR..AFYTTGE..I.I.GDIRQ.AHC NLSRAKWNN.TLKQIVKKL..RE..QFG....NK...T.IVF.NQS.SGGDP.EIVMHSF.NCGG.EFFYCNT
B_TH7.99 --S----V---------- ----------.---L..---Q..-W----Q..-.-.-.-----.--- ---ST--G-.--R--AE--.--..---...--...-.---.---.-----.---R--V.----.-----D-
B_TT.QZ4589 ---------N-----L-------V-- ----------.----..----..---.---..-.-.-----.--- ----T---D.------I--..--..---E...--...-.---.---.-----.-------.----.------S
B_TW1.138 --------KK--K-- -----RHI-R.R---..-----..S-V.--N..H.N....-K.--- ----SE-DD.--A---G-F..-
B_TW1.236 -P--L-----S-----KNP-K-- ----G----R.---M..-L--..------G..-.-.-----.--- ----TE---.--TE-A--F..K-
B_TW1.237 D--S--T--------K--K-- ----------.----..----..------D..-.-.-----.--- ----TE--E.--RK--E--..-K..-
B_TW1.257 ---N----------DT-K-- ----------.-V-M..---S..TL-A--A..-.-.-N---.--- ----IN-ED.---K--I--..K-..-YK
B_TW1.261 --K---V-N---------K---I-- ------TKIR.H---..-----..T--AR.-..K.-.-----.--- --NKTE---.--VK--E--..--
B_TW1.358 -----V-N---------P-- -I--------.R-TL..--R-..V------..-.-.----R.--- -I-E--XS-.------N--..K-..--K....--T..-
B_TW1.366 -------I-M----K--T-- ----------.----..---K..------.D..-.-.----K.--- -I-G-E--H.---L--E--..-K..---E...--...-.---.--
B_TW1.382 ----------RKP---- ----------.----..----..---A-.D..V.-.------.--- -I--T---E.----------..--..K-K....---...-
B_TW1.384 --A----------K-T-N-T ----------.--PM..---K..---A--D..-.-.------.--- -I-KTA---.------E--..--..--K....--...-.---
B_TW1.386 ---I---------T-- ----------.----..---K..------.-..-.-----R.--- -INKTT--H.--EL--E--..--..---E...--...-
B_TW1.391 ----------K--T-H ----------.--S-..----..------.N..-.-.----K.--- -I---D--D.--T---G--..--..-Y-K...-N...S
B_TW1.394 --L---N---V------Q--V-- ---------R.--P-..----..-----.D..-.-.----K.--- -I--KD---.------G--..--..---....--...S
B_TW1.646 -----K--T-- ---LHKI---.R-IL..-----..S-----T....-.E-K-K.--- -I--TD--H.---L--E--..--..-Y-
B_TW1.84 See inserts
B_TW1.90 See inserts
B_TW1.91 -----V-------V------K--T-I -----KI---.G---..----..------T....-.E---K.--- -I--TN--H.------G--..--..-Y-E...--...-.---
B_US.ACP ---------N---I------K--K-- -I--------.G-G-..----..TV--AEK..-.-.------.--- -I--TE--V.--Q-------..-D..---N...--...-.---.---.-----.-------.----.-----K
B_US.ADA -----S-------N-----K------ ----------.----..----..--------..-.-.------.--- -I--T----.--N--AT--..K-..---N...--...-.---.---.-----.-------.----.------
B_US.ALA -----D---------------------- -----IY.--GR---..----..--H--RQ..-.-.EN---.--- -I-----ED.------E--..--..--K....--...-.---.---.-----.---T---.----.------
B_US.BAL -----A--A----V---------- ----------.----..----..--------..-.-.------.--- --------D.--NK--I--..--..---...--...-.---.KH-.-----.---T---.----.------
B_US.BCSG3 -----A--S----------K-A---- -----KK---.R-TT..----..VY-----..-.V.------.--- --------D.--R---I--..--..--E....--...-.---.---.-----.-------.----.-----KS
B_US.BRVA ---------N-V---------- ----------.R-TM..-----..VY----Q..-.-.----R.--- ----S--E-.------T---..-V..--K....--...-.---.-R-.-----.-------.----.---F--
B_US.BWB ---------------T---------- -I--------.G-NV..----..-L----D..-.-.------.--- -I------K.--Q---T--..--..--V....--...-.---.-K-.----L.-------.----.------S
B_US.CDC42 ---------N----------V----- ------H---.RVTL..----..VW------..-.L.-N---.--- -I---Q---.--Q--ATT-..--..---...--...-.-A-.---.-----.-------.----.------
B_US.Diaz -----D---N---P-----KD--Q-- ---------R.--N-..----..---A---..-.-.------.--- -----A---.--A------..--..--K
B_US.JFL --------M----------K-IA-- ----------.--TL..----..------D..-.-.------.--- ----VQ---.------I--..--..---...--...-.---.K--.-----.-----T-.----.------
B_US.JM --V--D---N----------KT-V-- ----------.-----..A---..------D..-.-.----K.--- -I-KEN-Y-.----VAE--..G-..--K....-R...-.---.-H-.-----.---R--L.IVE-..I---D
B_US.JRCSF -----D-----------------K-- ----S-----.----..----..--------..-.-.------.--- -I---Q---.------E--..--..--N....--...-.---.TH-.-----.-------.----.-------
B_US.MA ---------E-H-I-----K-P---- ----G-----.--PM..----..---A--D..-.-.----K.--- -I--------.--E-V-----..--..---...--...-.-I-.---.-----.-------.----.------
B_US.MN ----------------H-----Q-- -----Y-K--.R---..----..-----KN..-.-.-T---.--- -I------D.--R---S--..K-..--K....--...-.---.---.-----.-------.----.------
B_US.NY5CG ---------N-V---------- ----------.G-A-..-----..TL-AREK..-.-.------.--- --------D.------T---..K-..--R....--...-.---.---.-----.-------.----.-----K.
B_US.P896 I----------------------V-- ---------R.RLS-..----..---ARRN..-.-.------.--- -I--------.--Q---I--..--..K-R....--...-.-
B_US.PD See inserts
B_US.RJS See inserts
B_US.SBA -----D---N---------K-A---- ---------R.--N-..----..------D..-.-.------.--- ----T----.----RGE--..--..--K....--...-.---.---.-----.-------.----.------
B_US.SBB ---------N----------T---- --------S-.R-S-..----..--VAAR-..-.-.----K.--- -I-IT----.--RL--S--..Q-..---K...--...-.---.-H-.-----.---T-T-.----.------
B_US.SBC -----D---N----------Q-- -I---------.--S-..----..--------..-.-.------.--- -----N-A-.---K--E--..--..--K....--...-.---.---.-----.-----N-.----.------
B_US.SC --L----------K-A---- --------TR.----..----..---A--D..-.-.------.--- -I--------.------I--..-D..--E....--...-.-I-.-R-.-----.-------.----.------
B_US.SF128 -----D---N----------V-- ----------.--N-..-----..-I----A..-.-.------.--. T-NKTQ-D-.--R--AI--..--..--K....--...-.-S-.K--.-----.----Q--.----.--------
B_US.SF162 ----------K------ ----------.--T-..----..---A--D..-.-.------.--- -I-GE----.------T---.QA..---..---...-.---.K--.-----.-------.----.------
B_US.SF2 -----D---N----------A-- ----------.--Y-..----..--H---R..-.-.----K.--- -I---Q---.--E------..--..---N...--...-.---.---.-----.-------.--R-.------
B_US.SF33 -----D---N-----L----V----- -------R-R.R-TS..---K..VL------..-.-.----K.-Y- -I------K.--E-VAT--..--..---...--...-.---.K--.-----.-------.--R-.------
B_US.TN100 -V----------IQ-- ---------R.--N-..-----..-W------..-.-.------.--- -----A-K-.---
B_US.TN1000 -------I---------- ----------.--S-..----..------D..-.-.------.--- -I-
B_US.TN1002 I------I-N---I---------- ----------.----..----..---A--D..-.-.------.--- -I-K-T-
B_US.TN1003 -----D-I---T-------K-P---- ----S-----.----..----..-------A..-.-.------.--- -I-K---S
B_US.TN1005 ---------N---I-----K------ -I--------.----..A---..---A--D..-.-.-N---.--- -I-EE---K
B_US.TN1006 D----A--S----N-----KDP---- ----------.--PM..----..---A--D..-.-.------.--- ---GTN-
B_US.TN1007 ----------G-------K---V-- ----------.----..----..--------..-.-.------.--- -I--E--
B_US.TN1009 ---------N----------IV-- ---------R.--N-..----..---A--Q..-.-.------.--- -I-A
B_US.WM ---------N---------K-P-Q-- -I--------.G---..---K..--------..-.-.-N---.--- -I--VT---.----V----..--..--K....--...-.---.K--.-----.-------.----.---F--
B_US.WMJ22 I----------H-------- ----Y--V-R.-LS-..----..--R-R.-..-.-I---.--- -I-------.------E--..--..--K....--...-.---.-H-.-----.---T---.----.------
B_US.WR27 -----A---N---------K------ ----G-KI-R.R---..----..---.-.D..R.V.------.-Y- -I-GT--K-.--EK--A-I..--IK--K....--...-.---.-H-.-----.-------.----.------S
B_US.YU I---------N----------V-- ----------.--N-..----..-L-----..-.-.------.--- ---KTQ-E-.--E--AI--..K-..---N...--...-.-I-.KP-.-----.---T---.----.------
B_US.twinABcon I-SK----SN-V---M---T--IH-- ----S-----.----..----..---A--D..-.-.------.--- -I--E---V.--QKV-E--..G-..--K....--...-.---.L----.-----T-.----.------S
B_US1.HC0 ----A-------I------A----- ----------.G---..----..---A---..-.-.------.--- -I--E----.--R-V-T--..--..---...--...-.-I-.-H-.-----.-------.----.------
B_US1.HC05 --------N---I---H--KT-N-T ---------R.--PM..---K..-----.-..-.-.-N---.--- ---K-E---.--R------..--..--K....--...-.---.-H-.-----.------
B_US1.HC06 ----A-----T-I---------- ----S-Y-G-.RLS-..----..--VA-RK..-.-.------.--- -I------D.---R--I--..--..---...--...-.---.D-P.------.---K---


V3 LOOP
B_CONSENSUS_96 VVIRSENFTDNAKTIIVQLNESVEIN CTRPNNNTRK.SIHI..GPGR..AFYTTGE..I.I.GDIRQ.AHC NLSRAKWNN.TLKQIVKKL..RE..QFG....NK...T.IVF.NQS.SGGDP.EIVMHSF.NCGG.EFFYCNT
B_US1.HC07 ----------T---------T--T-- ----G-----.--TM..---K..V--A.--..-.-.-----.--- -----A--D.------G--..Q-..---.....---...-.---.-H-.-----.--------.VE-N.F
B_US1.HC08 --------K---N-----K------ ----------.--PM..----..-------..-.-.----K.-Y- TVNKT-----.--AK-AS--..--..--E....--...-.---.TP-.-----.--A--N-.--R-
B_US1.HC09 ----A--S-----------K--N-T -V--S----Q.--P-..---K..---A---..-.-.----K.--- ---E-E---.--T------..K-..--N....--...-.---.---.-----.--------.-----.K-
B_US1.HC1 --------N---------K---K-- -I-------R.--NM..----..------D..-.-.-----.--- -I-K-E---.------Q--..K-..--R....--...-.---.---.------.-V-T---
B_US1.HC10 ----------------KN--G-- ---------R.-V--..---S..-L---.D..-.-.-----.--- -----N--K.--E---T---..G-..---N....--...-.-A-.-S-.-----.---F---.SF
B_US1.HC12 --------N-S-------K---V-- ---------R.----..----..-W----K..-.T.-----.--- TIN---X-E.-----AI-F..--..--K....GX...-.---.-H-.-----.--------
B_US1.HC13 -------S----------K------ ----------.R-TM..-----..VL----Q..-.-.--V-R.--- -I-GVQ--K.A---V-A---..K-..--R....--...-.---.KPP.-----.---T---
B_US1.HC16 ----------V----------Q-- ----------.----..A---..---A---..-.-.-----.--- ---SI---D.--R--A----..K-..----....--...-.-I-.---.-----.--PEI
B_US1.HC18 --------N----------V----- ----------.G--..----..---A--D..-.-.-----.-Y- -I--------.----V-T--..--..---....IT...-.M--.YH-.-----.PV--L
B_US1.HC19 ----------------K-P---- ----S-----.G-P-..----..------G..-.-.-----.--- --------D.------T--..K-..--E....--...-.---.---.-----.--------.----.--.
B_US1.HC2 --------N---------K---K-- ---------R.--NM..----..------D..-.-.-----.--- -I-K-D---.--E-------.E-..--K....--...-.-A-.---.-----.-V-----.----.------
B_US1.HC20 -------KDP---- ----------.----..A---..---A--G..-.-.-----.--- -I--------.--G---S--..--..H--....-TT..-.VA-.QS
B_US1.HC22 ----D----------K----- ----S-----.--T-..----..--------..V.-.-----.--- -I--T---D.------R--..G-..K--.....-E...-K-I-.EP-.-----.--TT---.-----.D-STVI
B_US1.HC23 ----D-----V---------A-V-- ---------R.G---..----..--------..-.-.-----.--- -I--------.------V--..-V..---....--...-.---.-H-.-----.--------.----.----
B_US1.HC24 ----------------------I--- ----------.--P-..----..---A--D..-.-.-----.--- -----T---.------T--..--..---....-R...-.---.---.-----.--------.----.----
B_US1.HC25 I-----L----------K-P-I-- ----------.-----..---K..--DA-.D..-.-.-----.--- -V-STE---.--REVAT--..KG..L--.....-T...-.VA-.---.-----.D--I---.----.----
B_US1.HC26 ---I---------T---- ----------.----..----..-------..-.-.----K.--- -I---Q---.--ER--I---..G-..---....KTN..-.---.---.------.------V.----.----
B_US1.HC27 N---I---------V-- ----------.G--M..----..------A..-.-.----E.--- -I--K--E-.--G--AS--..--..--K....-TT.T-.L-
B_US1.HC28 ----D----------V------ ---------R.--T-..----..---A-.D..-.-.-----.--- -I-GTR---.--E-
B_US1.HC29 --------N-----L----A--A-- ---LS-K--R.----..----..---A-.D..-.-.-----.--- -I--ED---.------I--..G-..---....-TT..-.-A-.-K-.-----.--------.----.----
B_US1.HC3 ----D--S---R----------V-- --------SR.R-S-..----..--TAREG..-.-.-----.--- -I-G-E-ES.---R--E---..G-..--K....--...-.---.-H-.-----.--------.----.----
B_US1.HC30 ------IS----------V-- ---------R.-----..A---..---A--D..-.-.------.-Y- TINKT--H-.----V-E---..--..--N....-T...-.---..R-.---
B_US1.HC32 ----------------Q----- ---------R.-----..---K..---A--G..-.-.-----.--- ----TA----.---L--D---..--..----....--...-.---.-----.---T---A----.-----
B_US1.HC33 ----------T------ ----------.----..----..------D..-.-.-----.--- -I--T--E-.-----AT---..--..-Y-N...-T...-.---.---.-----.---T---.----.-----
B_US1.HC34 ---------T--------KDP---- ----------.----..----..---A--D..-.-.-----.--- -I--------.--R---DIC..-V..---....--...-.---.---.-----.-----T-.----.-----
B_US1.HC35 --K----------P---- ----------.----..----..------D..-.-.-----.--- ----------.------V---..--..----....--...-.---.-----.-----T-.----.-----
B_US1.HC36 ----D-------I---------V-- ----------.----..---S..------D..-.-.-H---.--- --------D.------I---..--..--E....-R...-.---.K---.-----.---T---.----.-----
B_US1.HC37 --------N--R----H----IV-- --------G-.---L..A----.G-HA---..-.-.-N---.--- ---AEE-D-.--EK-A----..G-..R-P....--...-.-N-.TK-.-----.--------.----.----
B_US1.HC38 ------S-N---N-L---KAPI--- ----------.G-A-..----..TV-A--R..-.-.-----.--- ---ST--D-.--G--AI---..--..----....--...-.-T-.---.-----.--K-----.----.----
B_US1.HC39 ----D----------QT---- ----------.----..----..---A--G..-.-.-E---.--- ----V--E-.--E--AI---..--..--K....--...-.---.---.-----.--------.----.----
B_US1.HC4 ----------------T-P---- ----------.----..A---..---A---..-.-.-----.--- TI------D.--Q--AI---..S-..---....-N...-
B_US1.HC40 --------N----------A-- ----------.G-P-..----..------D..-.-.-----.--- -----D-D-.---L--T---..--..----....-E...-.-I-.---.-----.--------.----.----
B_US1.HC42 ----K----------------A-- -S---------.G---..----..---A--D..-.-.-----.--- -----D---.---L--EN-..G-..HL-....MQ
B_US1.HC43 ---------L-N-----K-P---- ----G----R.----..----..-------..-.-.-N--L.--- -I--------.-----AD--..--..--E....--...-.---.-H-.-----.-F-T---.----.-----
B_US1.HC44 --------N----------V-- ----------.--P-..----..---A--D..-.-.-----.--- ----EA----.----V-A---..--..---K...G-....-.-I-.-H-.-----.-----T-.----.-----
B_US1.HC45 --------N----------T---- ----------.----..----..------D..-.-.-----.--- -----Q---.------E---..--..--K....-S...-.---.KP-.-----.---T---.----.-----
B_US1.HC46 -.L--D----------------V-- ----------.G---..----..---A---..-.-.-N---.--- -I-A-N-T-.-----AI---..-K..---....-N...-.-I-.-R-.-----.---T---.----.-----
B_US1.MA145con E-H-V-----K-P---- ----G-----.--PM..----..-WFA--D..-.-.----K.--- -I--------.--E-V----..G-..--E....--...-.-I-.---.-----.--------.----.-------
B_US10.SFBUcon See inserts
B_US10.SFPEcon ----------------------IA-- ----------.----..----..---A-.D..-.-.-N---.--- -I--KD---.---L-----..--..--KEQLV--...-.---.---.-------.----.------S
B_US11.074 I------I-N---------K-P---- ----------.--N-..----..-------..-.-.-N---.--- -INGT-----.------E---..--..---....--...-.-I-.---.----L.--------.----.------S
B_US11.107 ---------N------I----T---- ---------R.G-N-..----..---A-.G..-.-.----K.--- -I-GT-----.------Q---..--..---....--...-.---.---.L-----.---T---.--T-.-------
B_US11.306 I------L-N---------K-P---- ---------R.G----..---Q..--------..EVT.----R.-Y- -I-S----E.------IQ-..-K..---....--...-.-S-.---.T---I.---T---.----.-----
B_US11.333 ---------N---I---------- ----------.----..----..-------..-.-.-----.--- ---------D-.-----AI---..-D..---V....--...-.I.-I-.---.-----.--------.----.-----DS
B_US11.349 -----D----------V------ ----------.-VTM..----..VW-----..-.V.----K.--- ----TT----.--R-V-E---..--..-K-R....-AT..-.---.-H-.-----L.--------.T---.--------
B_US11.419 ---------------------- ----------.----..---Q..-L-A--D..-.-.-----.--- -I--------.------E---..KV..---....--...-.-I-.---.-----.--------.----.--------
B_US11.44 -I-----------------A-I--- ----------.---L..-----..SV----K..-.V.-----.--- -I-G------.A-----E---..-D..--K....--...-.---.-K-.-----.---T---.--A-.-------
B_US11.550 ----------------K------ ----------.----..----..-------..-.-.-----.--- ----E--EK.--D---A---..S-..---....--...-
B_US11.552 ---------N---I---------- -I---------.---P-..----..--FA--D..-.-.-----.--- -I---N----.-----T-N-..KD..--R....X-...-.---.----.T-----.-----N-.-
B_US11.725 -----D--S----------TP---- ----------.--P-..----..------D..-.-.-----.--- -I-KDD----.------TN-..-D..---....--...-.---.-R-.-----.--------.----.------S
B_US11.UK --------S----------K---K-H -I---------.---M..---K..-Y-A--D..-.-.-----.--- -I-E-H--R.--G---T---..--..---....--...A.---.---.-----.---T---.----.------S
B_US12.CHB2A -----N----------T---- ----G-----.--N-..----..-W-A-.D..-.-.-----.--- -I--------.--Q--A---
B_US12.CHBMOM ---------I---H-------- ----------.R-S-..----..--R---N..-.-.-----.--- -I--------.--EKVA---
B_US12.LC1A -----A----------A---K --S-------.----..----..--DA--A..-.-.------.-Y- TR-GTR---.--EKV-E-I
B_US12.LC2A ----------------------V-- --------KR.--N-..----..---A---..-.-.-----.--- ---TT---D.--G---
B_US12.LC3A -D--S----------K----- ----S-----.G-N-..----..-V------..K.-.-----.--- -----R--D.------E--
B_US12.LC4A ----A----------K-R---- ---------R.---M..---K..---G-.D..-.-.-----.--- -----A---.------E--
B_US13.141con -I------S----------I-- ----S-----.--P-..----..---A---..-.-.-----.-Y- -I--KE--K.--R---P--..--..-V-V...--...-.-I-.-S-.-----.---Y---.----.-------
B_US13.144con -----D-----V-------K-A---- ----S-----.G--L..----..-I---EQ..-.T.-----.--- -I--E-----.------G--..--..----....-AT..-.---.---.-----.---T---.----.-------
B_US13.149con --------SN---N----------T -I--G----R.---L..Q---..---SS-G..-.-.----K.--- TINIPQ---.---LV-D--..-N..--K....-R...-.-A-.T-PPP-----.-V--Y--.----.-------
B_US14. --V--A---N---I---------Q-- -I---------.----..----..-I-A--G..-.-.----R.-Y- -I--T----.-----T---..Q-..---....K-...-.---.K--.-----.-V-----.----.-------


V3 LOOP
B_CONSENSUS_96 VVIRSENFTDNAKTIIVQLNESVEIN CTRPNNNTRK.SIHI..GPGR..AFYTTGE..I.I.GDIRQ.AHC NLSRAKWNN.TLKQIVKKL..RE..QFG....NK...T.IVF.NQS.SGGDP.EIVMHSF.NCGG.EFFYCNT
B_US14.2 -----V---N--R------R-A---- ----S-----.----..----..------N..-.-.-----.--- -I--G----.------S--..--..---.....--...-.---.---.------.-------.----.------S
B_US14.3 ---------------------------- ----------.--P-..----..---A--D..-.-.-----.--- -V-K----S.--E-------.G-..---.....-N....-.---.---.------.-------.----.------S
B_US14.4 I-L---------------------- -I---------.----..----..---A--D..-.-.-----.--- YI-K-N-T-.--E---E--..--..---N...--...-.-I-.-S-.-----.---F---.----.-------
B_US15.A --------N---S---------- -I---------.--NM..-----..------D..-.-.-----.--- -I-E------.------N--..--..--R....---...-.-I-.---
B_US15.B ---------N--RI----------H ---------R.--NM..-----..------D..-.-.-----.--- -I--------.-----AN---..--..--E....---...-.-I-.---
B_US15.C I----V-L-----I---H------M- ----------.----..----..-------..-.-.-----.--- ---S----D.---N-AI--..--..K-E....--...-.---.---
B_US15.D ---S-A-L---------K--D--VM- ----S-----.--PM..-----..------D..-.-.-----.--- -----D---.---RVAI--..--..--.....---...-.-I-.K--
B_US15.E I----A--SE-D---------T--M- ----------.RMT-..-----..V------..-.-.-----.--- ---E------.--R--AI---..K-..--E....---...-.---.-H-
B_US15.F -----A-L---------------A-H -I--------R.--P-..----..---A--D..-.-.-----.--- ----EE---.---------..--..--K....---...-.---.---
B_US16.NYRT ---------N---I---H-------- ----------.G---..--WG..T-FA--N..-.-.-----.-Y- -I-----D-.---K--D--..--..---.....---...-.---.-----.-------.--R-.-------
B_US16.NYRT3 -----A----------K-r--- ----S---S-.G---..-----..-----EA..-.T.----R.-Y- -I---A--E.--G---E--..--..--E....--...-.-A-.-K-.----L.-------.----.-------
B_US17.CB -----A-L----------K-Y---- ----------.--S-..----..-------..-.-.-----.--- -I-GT----.--S---T--..--..E-K....--...-.-I-.---.-----.-------.----.-----D-
B_US17.CB2 --------M--V------K---K-- ----------.G-N-..----..---A---..-.V.-----.--- ---S-E--D.A-N-T-T--..--..H-K....--...-.---.QP-.----L.---T---.----.-------
B_US17.CB3 I--------N-V-N--------I--- ----H-----.---M..---G..-L-A--R..-.-.-N---.--- -I--------.--EK--R--..--..K-...--...-.-N-.T--.----R.-------.----.G-L---S
B_US17.CB4 -----D--S----------KDP---- --------I-.G---..----..---A--Q..-.V.-----.--- -I--T---K.--H--AT--..--..--K....-IT..-.-S-.---.-----.-------.----.------S
B_US17.CB5 -------------L----K-P---- ----------.--N-..----..------Q..-.-.-----.--- -I-------.---G-------..--..--R....---...-.---.---.-----.-------.----.-----S-
B_US17.CB6 -----D----------K-I---- ---------R.--N-..----..------D..-.-.----R.--- ----T---D.-----AN--..--..--.....-AT..-.-.---.-H-.-----.---T---.----.-----DS
B_US17.CB8 I----------V-----HK---I--- ----------.----..----..------Q..V.-.-----.--- ---E-----.-----AI--..--..--V....--...-.---.-K-.-----.-------.----.-----DS
B_US17.CB9 -----D---N----------A---- ----------.----..----..---A--G..-.-.-----.--- V-DK-----.------R--..--..--Q....---...-.---.---.-----.-------.----.-------
B_US18.LBAcon -----T-I---------- ----------.--N-..----..------D..-.-.-----.--- -I-EK--F-.----VAG--..--..--K....---...-.---.--P
B_US18.LBB ---N----------T-Q-- -S--------.G---..---S..---A---..-.-.-----.--- -I--V----.--GY-----..--..---.....--...-.---.KT-
B_US18.LBBB --S---------H--K--K-- --------S-.G---..---G..--F---R..-.-.-N--H.--- -I--KE---.--GL-TE--..ED..--K....I-...-.VT-.-H-
B_US18.LBD ---N-------H---T-N-T -V-------R.----..----..---AA-D..-.-.-----.--- -I-G-E-T-.---KV-T--..Q-..----N...--...-.-I-.RP-
B_US18.LBE ---N-----T---T--IK-- ----T-----.----..E---..--------..V.-.-N---.-Y- -I-K-D--K.--AKVAE--..--..-YK....--...-.---.K-P
B_US18.LBF -----V--------T--Q-- ----G-----.---M..---K..V--A.--..-.-.-----.--- -I--KE--D.---EKV-N--..-N..-W.....-N....-.-I-.S
B_US18.LBG ----------KT-K-- ----------.----..----..---A--D..-.-.----K.-Y- ----TE--T.--G-V-E--..--..--
B_US18.LBI ----------K-A---- ----S-----.---L..----..-W-----..-.-.-----.--- ---------.------E--..--..R-E....--...-.---.-H-
B_US18.LBJ --S----I-----K-P---- ----------.----..----..------D..-.-.-----.--- -I--------.--S---AI--..--..---.....--...-.---.-K-
B_US18.LBK -------I------A----- ----H-----.----..----..-------..-.-.-----.--- --------D.-------I--..--..---.....-TT..-.---.-R-
B_US18.LBLCN ---G-------H-----Q-- ----S-----.----..E---..---A-.D..-.-.-----.--- -I--TH---.---K-----..--..--.....--...-.---.QP-
B_US18.LBLCT ----------K-PIQ-- ----------.-VP-..----..---A---..-.-.-----.--- ---SVT---.---------..--..---.....--...S.-A-.---
B_US18.LBM ---N----------T---- ----------.----..----..--F---D..-.-.-----.--- -I---R--E.--NR--T--..--..---N...--...-.---.-H-
B_US18.LBO1 ------EI------KP---- --------SR.---L..-----.TL---.-..-.-----.--- --------D.--T---T--..--..-Y-....--...-.---.---
B_US18.LBPATH2 --S------VI-----A-- ----------.G---..----..--------..-.-.-----.--- -I-LG----.--N---T--..-G-....--...-.---.---
B_US18.LBPTCC1 ----------K---V-- ----------.----..----..-------..-.-.-----.--- ----TQ---.--T-------..--..----.....--...-.---.-H-
B_US18.LBQ -------I------K-ID-- ----S-----.----..----..------D..-.-.-----.--- -I------D.------T--..--..---.....--...-.---.---
B_US18.LBSDIR ---N----------- ----------.--S-..----..------D..-.-.-----.--- -I------D.---R--I---..-K..-YK....--...-.---.---
B_US18.LBV --S----------K----H -I--------.R-TL..-----..VY-----..-.-.-N---.--- -I---N-TD.------T--..K-..----.....--...-.---.KP-
B_US18.LBW ---N----------A-V-- ----------.G---..----..---A--D..-.-.-----.--- ---KVA---E.---KV-E--..--..--K....K-...I.---.-S-
B_US18.LBY ---N-I-N----------H -I--------.G---..A----..T--A--A..-.-.-----.--- ---I-N-T-.---R------..-D.----E....--...-.---.-H-
B_US18.LBZ -L---------H--K--V-- --------I-.--R-..-----..-W-----..-.T.-----.--- -----D---.--R-V-M--..--..H-K....--...-.---.---
B_US19.VA1con -- ----G----R.---M..---Q..-L-----..V.V.-----.--- -I--VD---.----V-T--..--..-
B_US19.VA2 -- ----SI-K-R.-VR-..---G..-W-K--Q..V.T.----R.--- -----D--D.------N--..--..-
B_US2.D2US656 ----------V-- ----------.----..----..---A--D..-.-.-----.--- ----TQ---.------I--..--..---N...--...-.---.---.-----.--------.-
B_US2.D2US657 ---N----------TIK-- ----------.G---..----..--------..V.-.-N---.--- -I-STD-GK.--EKV-E--..-K..---S...--...-.---.-H-.-----.-------.----.------
B_US2.D2US660 -L---T-------KNP-V-- ----------.--PM..---K..-M-A---..-.-.----K.-Y- -I--TE---.----------..EG..--K....--...-.---.DR-.AD---.---L-T-.----.-------
B_US2.D2US662 ---N---------K-A---- ----------.----..----..--------..-.V.-----.--- -I-ST----.--G------..--..--D....-R...-.---.DK-.------.----LN-.----.------S
B_US2.D2US664 -------I-V-H--K--V-- ----------.----..----..---A-.D..-.-.-----.--- ---K-D-EE.A-N-TAR--..K-..-L.....-TT..-.---.---.-----.-------.----.--I---S
B_US2.D2US665 -I---------H----I--- ----------.--PM..----..--------..V.-.-N---.-Y- -I-K----D.------T--..--..H-K....--...-.---.---.-----.-------.----.-----D-
B_US2.D2US704 ---N----------T-V-- ----------.----..----..--H----..V.-.-----.--- -I-GE----.--EK------..K-..--R....D-...-.---.KH-.------.-------.----.--Y----
B_US2.D2US705 ---N---------K-A-K-- ----------.--N-..----..---A---..-.-.-----.--- -I--TD---.--GK--E--..-A..--.....-CS..K.---.K--.-----.-------.----.-------
B_US2.D2US706 ----------------I--- ----------.--N-..----..--------..-.-.-----.--- ---KE--ET.-----AE--..-K..-Y.....--...-.---.K--.-----.------
B_US2.D2US714 --S----N-L-------N-T -I--------R.---M..----..---A--D..-.-.-----.--- -I--T---E.--QR------.G-..--.....--...-P---.---.-----.-.
B_US2.D2US717 ---------------P-KMD -I-------R.--N-..----..------D..-.-.-N---.--- -----N-T-.--QK-AI--..G-..--Q....--...-.---.---.-----.--T-I--.----.-----SS
B_US2.D2US724 --SN--A-------A----- ----G-K-IR.--SM..----..--R.--Q..-.-.-N---.-N- T-------E.--R-V-----.SK..---.....--...-.-A-.K-P.-----.---T---.----.-------
B_US2.D2US725 ------------------K------ ----------.--P-..----..---A--D..-.-.-----.--- -I--------.------E---.K-..----.....--...-.---.-----.--
B_US2.D2US726 --------------KP-V-- ---------R.-----..A---..--HA--N..-.-.-----.--- ----TQ-D-.--H---T--..-K..---A...D-...-.---.---.-----.-------.----.-------
B_US2.D2US727 ---N----------A-- ----S-----.-V--..----..--------..-.-.-----.--- -I-AT----.-------D--..K-..----.....--...-.---.K--.-----.-------.----.-------
B_US20.C1con -----A-LS--S-I---------VM- ----------.--N-..----..--------..-.-.-----.--- -----N--D.--EK--I--..--..--Q....--...-.---.---.-----.--T-I--.----.------S
B_US20.C2con ---------N---------K-P---- ---------R.-----..---S..-L-----..-.-.----K.-Y- TIDA-----.---------..--..-K....-N...-.V--.Q-A.-----.---T---.----.-------
B_US20.C3con ---------N---------KDA-K-- -S--G----R.R-S-..-----..--VA-KR..-.-.-----.--- -I-KKD---.--N-V----..QG..--.....--T..-.-I-.QPH.-----.---T---.I---.-------

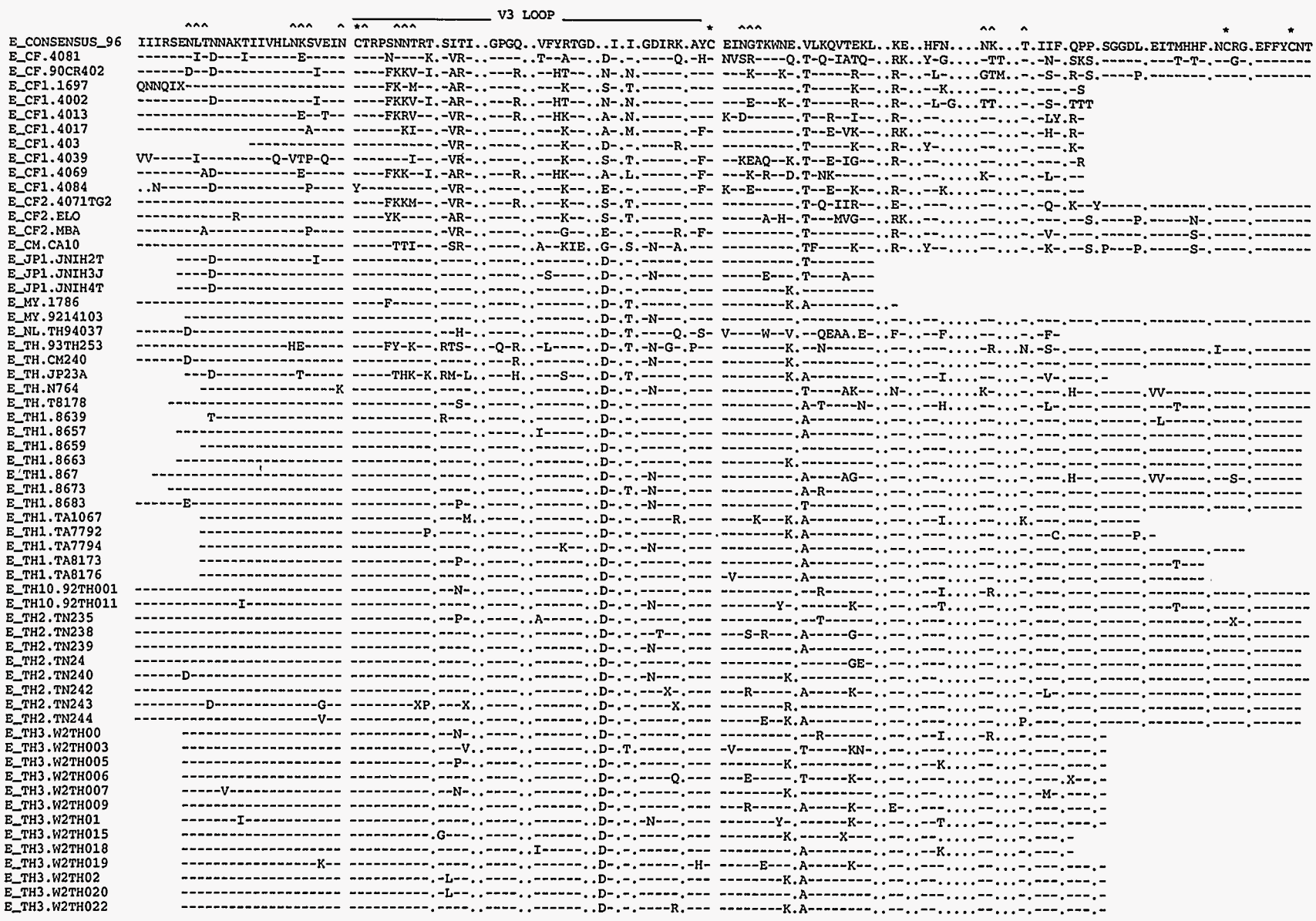
V3 LOOP
B_CONSENSUS_96 VVIRSENFTDNAKTIIVQLNESVEIN CTRPNNNTRK.SIHI..GPGR..AFYTTGE..I.I.GDIRQ.AHC NLSRAKWNN.TLKQIVKKL..RE..QFG....NK...T.IVF.NQS.SGGDP.EIVMHSF.NCGG.EFFYCNT
B_US21.P1
B_US21.P2
B_US21.P3
B_US21.P4
B_US21.P5
B_US21.P6
B_US22.E0105
B_US22.E0204
B_US22.E03A03
B_US22.E04B03
B_US22.E0502
B_US22.E0607
B_US22.E0704
B_US22.E0803
B_US22.E0903
B_US22.E1001
B_US22.E11A06
B_US22.E12A03
B_US23.201
B_US23.303
B_US23.811
B_US24.85W3F
B_US24.85W6A
B_US24.8674A
B_US24.87Y3E
B_US24.88B4E
B_US3.D1US712
B_US3.D2US71
B_US3.D2US715
B_US3.D2US716
B_US4.ZhuPtA
B_US4.ZhuPtF
B_US4.ZhuPtL
B_US4.ZhuPtR
B_US4.ZhuPtV
B_US5.pt5bl
B_US5.pt6bl
B_US6.I0
B_US6.I02
B_US6.I03
B_US6.I04
B_US6.I05
B_US6.I06
B_US6.I07
B_US7.306
B_US7.565
B_US7.657
B_US8.R
B_US8.R2
B_US9.S
B_US9.S2
B_US9.S3
B_US9.S4
B_VE.VE
B_VE.VE2
B_VE.VE3
B_VE.VE4
B_VE.VE5
B_VE.VE6
B_VE.VE7

```
                                                  ___________________ V3 LOOP ___________________
                 ^^^           ^^^   ^ *^   ^^^                          * ^^^    ^^ ^                    ^^   ^     ^^^                  *      *
B_CONSENSUS_96   VVIRSENFTDNAKTIIVQLNESVEIN CTRPNNNTRK.SIHI..GPGR..AFYTTGE..I.I.GDIRQ.AHC NLSRAKWNN.TLKQIVKKL..RE..QFG....NK...T.IVF.NQS.SGGDP.EIVMHSF.NCGG.EFFYCNT
B_VE.VE8         -----A--S---R------------ ------Y-A-.R-S-..-----..--R-K-R..-.-.------.--- TINK----E.--GK--T---..--..--R....---.---.---.-----.-------.----.------S
B_VN1.HCM9       ---------N---I------------ ----------.--P-..----..---A----.V.-.------.-Y- -I--EXNKX.NQ
B_ZA.ZA504       -----D-L---------H------- -------IK-.VYR-..----..----I-K..-.T.-GM--.T-- -I-E----K.--E---I---..--..-L-N...--...-.---.---.------.--TT---.----.------S
B_ZA.ZA508       -----D---N---------------- ----------.----..-----..--------..-.-.--T--.--- -I-ST----.--R---G---..--..K--.---.---.---.---AN.-------.--R-.------L
B_ZA.ZA509       -----D-----------H-------- ----------.G-G-..-----..TV--A-K..-.-.-N---.--- -I--E---K.--E---V---..--..--RK...---...-.M--.K--.-----.---T----.----.------S
B_ZA.ZA510   See inserts
B_ZA.ZA512       ---K---------V-V---K-A---- ------F---.G-Y-..-Q--..-V----K..-.-.--KKK.-R- ---ET----.--N---E---..-V..---....--...-.---.---.-----.-------.----.------S
B_ZA.ZA513       I----D-----V-------H------ ---------R.C-Y-..----..--H.-....Y...-----.--- ---ST----.--R-VA---..--..-Y.....--...-.---.---.-----.--I--T-.----.------V
B_ZA.ZA524       I----D--S----------Q----- ----------.----..---S..---A--D..-.-.-----.--- -I--TQ---.---H-AI--..--..--R....--...-.-N-.TH-.-----.-----T-.----.------S
                                                  ___________________ V3 LOOP ___________________
                 ^^^           ^^^   ^ *^   ^^^                          * ^^^    ^^ ^                    ^^^       ^ ^^                 *      *
C_CONSENSUS_96   IIIRSENLTNNVKTIIVHLNESVEIV CTRPNNNTRK.SIRI..GPGQ..TFYATGD..I.I.GDIRQ.AHC NISKDKWNE.TLQRVGKKL..AE..HFP....NK...T.IKF.APS.SGGDL.EITTHSF.NCRG.EFFYCNT
C_BR.W2BR025     -----K---D----------G-N ----------.----..-----..A-----E..-.-.-----.--- ---RTA--K.---E-----..--..--L....--...A.---.-KH.-----.-------.----.------
C_BR1.HSP203     ----------A--------------- -I--S-----.----..-----..A-----G..-.-.-N---.--- -T--EN--K.---K---IS..QS..TSL....I-...-.---.DQH.-----.-------.----.-L-----
C_BR2.91BR015    -----K---D---------K-P---N ----------.----..-----..--------..-.-.-----.--- ---RTA-KK.---E-----..--..---....--...-.---.-NH.-----.-------.----.-------
C_BU1.91BU001    ----------A--------KDP---E ----------.----..-----..IL-----..-.-.-----.--- --NET---K.---D-SE--..-K..Y--....--...-.-N-.-QP.-----.--V----.----.-------
C_BU1.91BU002    ----------------Q-KDP---E -I-------E.----..----..----H-A..-.-.-----.-Y- ----AT--K.---E---E-..-K..---....--...-.-N-.TQP.-----.-------.----.-------
C_BU1.91BU003    ---------D-A----IQ-K-P---N ---------R.----..----..--------..-.-.-----.--- ---RG---K.---E-----..--..--Q....--...-.-N-.TK-.-----.---Q-I-.----.-------
C_BU1.91BU004    -----K--SD-A-V-----K-P--TE ---------E.----..-----..A------..-.-.-----.-Y- T-NGT---K.---E-----..--..Q-.....-SS..K.---.---.-----.-V-----.----.-------
C_BU1.91BU005    -K---K---D-A-I---------Q-N ---------T.----..-----..A-----E..-.-.-----.-Y- ---GE----.--K--KE--..E-..---....--...-.-Q-.-N-.-----.-------.----.-------
C_BU1.91BU006    -V---------A--------K----- --------KE.----..-----..--------..-.-.-----.--- ---GTN-SR.---E-SR--..--..---....--...A.---.---.-----.-VV----.----.-------
C_BU1.91BU007    VT-----M---I-I---Q-K-P---N ----------.G-G-..-----..--F--E-..-.-.-N---.--- ---GT---H.--EQ-MEE-..KK..---....--...-.-E-.---.-----.-------.----.-------
C_BU1.91BU008    V----K--S--A-I-----DK----T ---------T.----..-----..A-----E..-.-.-----.--- ----TA--D.---K-RT--..-G..Y-S....-A...-.-E-.TQP.A-----.-V-----.----.-------
C_BY1.BLR5A      ----------A-----Q-K-P---S ----------.----..-----..--------..-.-.-N---.--- ---R-G--K.--HN-TEE-..KK..---....--...-.-N-.-
C_BY1.BLR8A      ----------A-----Q-KNP---N ----------.----..-----..--------..-.-.-----.--- ---R-S--T.--HN-TVE-..KK..---....--...-.-N-
C_BY1.BLR9A      ---------D-A-------------- ----G-----.----..-----..A------..-.V.----K.--- ---RGD-TR.--R--RD--..QK..--.....--...S.-R-.D
C_CF1.15166      -------H-D-T-I---Q--K----- ----------.--SL..-----..A--------..-.-.-N---.--- -V--TD--K.M--N-T----.EK..I-.....--...N.-T-
C_CY.HO021       ---------D-A----------IK-- ----------.----..-----..--------..-.-.-N---.--- ---AEE--K.----------..G-..Y--....--...-.-E-.K--
C_DJ1.DJ259      ----------A-I---Q-----A-N ---------E.----..-----..--------..-.-.-----.--- ---RKE---.--RK-RE--..KT..L-R....TA...N.-T-.Q--.-----.-------.--G-.-----
C_DJ1.DJ373      ----------A-I---Q--Q----N ---------Q.----..-----..--------..-.-.-----.--- ---RQ--K-.---Q-KG--..K-..--.....--...-.---.---.----I.-------.--G-.-----
C_ET2.2220       T---F------A-I---Q-------T ----S----E.----..-----..--------..-.-.-----.--- ---EE---K.---K-KE--..QK..---....--...-.-E-.K--.-----.-------.--G-.-------
C_ET2.E1320                      KP---- ----------.-M--..-----..------A..-.-.-----.--- ---EKT--D.---E--R---.Q-..R
C_ET2.E1325                  I------KP---- ----------.-M--..-----..--------..-.-.-----.--- ---EKA---
C_ET2.E1439                        ---T ----------.----..-----..--------..-.-.-----.--- ---GE---T.--
C_ET2.E2564        ---------A-I---Q-K-P---T ----------.-M--..-----..------S..-.-.-----.--- ---EKT-SN.---E--
C_ET2.E6209                        ----N ----G----E.----..-----..--------E..-.-.-N---.--- ---RE----.--H--KE--..R-..-
C_ET2.E6613                    K-P-Q-N ---------E.----..-----..--------..-.-.-----.--- ---GS----.---E-TT--..G-..-
C_ET2.E72150                     P-Q-I ----------.-M--..-----..--------..-.-.-----.-Y- ---G-A--K.---D------.Q-..R
C_ET2.E7827                      --T---N ----------.----..-----..-L------..-.-.-----.--- ---GEN
C_GA1.G134       -----K---D-A-I---N--K----M ---------E.----..-----..--------..-.-.-----.--- ---GSQ-TK.---Q-A----..K-..KL.....--...-.-Y-.NA-.-----.-------.----.-------
C_GA1.LBV105     -----Q--S--G-N------------ ---------E.----..-----..A---P--..-.-.-----.--- ---ENQ-SK.--H--SE---..K-..Y--....--...-.-R-.NA-.-----.-------.----.-------
C_GM.GM3         ...------D-A----------P-E ----S----.----..-----..A-F------..-.-.-----.--- ----SV-KK.---M-----..--..---....-R...-.---.E--.------.-V---..........
C_IN1.D1044      --L----P-----------Q----M ----------.----..-----..--------..-.-.------.-Y- ---------.---M-----..--..---....--...-.-RS.---.-------.-------.----.-----
C_IN1.D747       ------------------Q----- Y---------.GV--..-----..--------..-.-.-----.--- ----H----.----------..--..---....--...-.-R-.-S-.-----
C_IN1.D757       --------A---------Q--R-- ----------.----..-----..--------..-.-.-----.--- ---EG----.----------..--..Y--....--...-.---.-S-.---G-
C_IN1.D760       ---------D---------Q---V- ----------.----..-----..--------..-.-.-----.--- ---E-----.----------..--..---....--...-.---.-A-.-----
C_IN2.D1024      ------------------Q----- ----------.----..-----..--------..-.-.----R.-Y- ---E-----.----------..--..---....--...-.-N-.-S-.-----.-------.----.-----
C_IN2.D744       ------------------Q----E ----S-----.----..-----..--------E..-.-.-----.--- ----GN-T-.A----S---..--..---....--...-.---.-S-.-----.-------.----.-----
C_IN2.D766       --------------L----Q----- ----------.-L--..-----..--------..-.-.----R.--- ---------.-----S---..K-..---....--...-.---.-S-.-----.-------.----.-----
C_IN2.D808       ------------------YQ----- ----------.-T--..-----..--------..-.-.-----.--- ---E-----.-----S---..--..---....--...-.---.-S-.-----.-------.----.-----
C_IN2.D868       ----------A-I-----DQP---- R---------.----..-----..--------..-.-.----R.-Y- ---EN----.-----S---..--..Y--....--...-.-Q-.-S-.-----.-------.----.-----
C_IN3.IND        ----------A--------Q----- ----------.----..-----..--------..-.-.----E.--- ---E-R--K.----------..--..R---....--...-.---.NS-.-----.-------.----.-------
C_IN3.IND2       ------------------K--D-- ----------.----..-----..--------..-.-.------.RNL ---R-L---.-FKM------.R-..-LL....--...-.---.-S-.-----.-------.----.-------
C_IN3.IND3       ----------------------E ----------.----..-----..--------..-.-.-----.--- ---------.-------R-..--..R---....--...-.---.-S-.-----.-------.----.-------
C_IN3.IND4       -V----D-KS--------------E ----S-----.----..-----..--------..-.-.-----.--- ----KA---.A----S---..--..---....--...-.-I-.N--.-----.-------.----.-------
C_IN3.IND5   See inserts
C_IN3.IND6       ---------E.---------Q----- -A---------.----..-----..--------..-.-.-----.--- --RE-----.-----S----.R-..---....--...-.---.TS-.-----.-------.----.----S--
C_IN3.IND7       ---------D---------Q----- ----------.----..-----..--------..-.-.-----.--- ---EA-----.-----R---..--..---....--...-.-N-.TS-.-----.-------.----.------Q
C_IN3.IND8       ------------------Q----E ----------.----..-----..--------..-.-.-----.--- ---------.---M-S---..--..---....--...-.-Q-.-S-.-----.-------.----.-------
C_IN4.CMCH1      ---------D---------------- ----------.----..-----..------E..-.-.-----.--- -------I-.----------..-
C_IN4.CMCH10     ------------------Q----- -A--S-----.----..-----..--------..-.-.-----.--- ---A-----.---K-S---..-
```

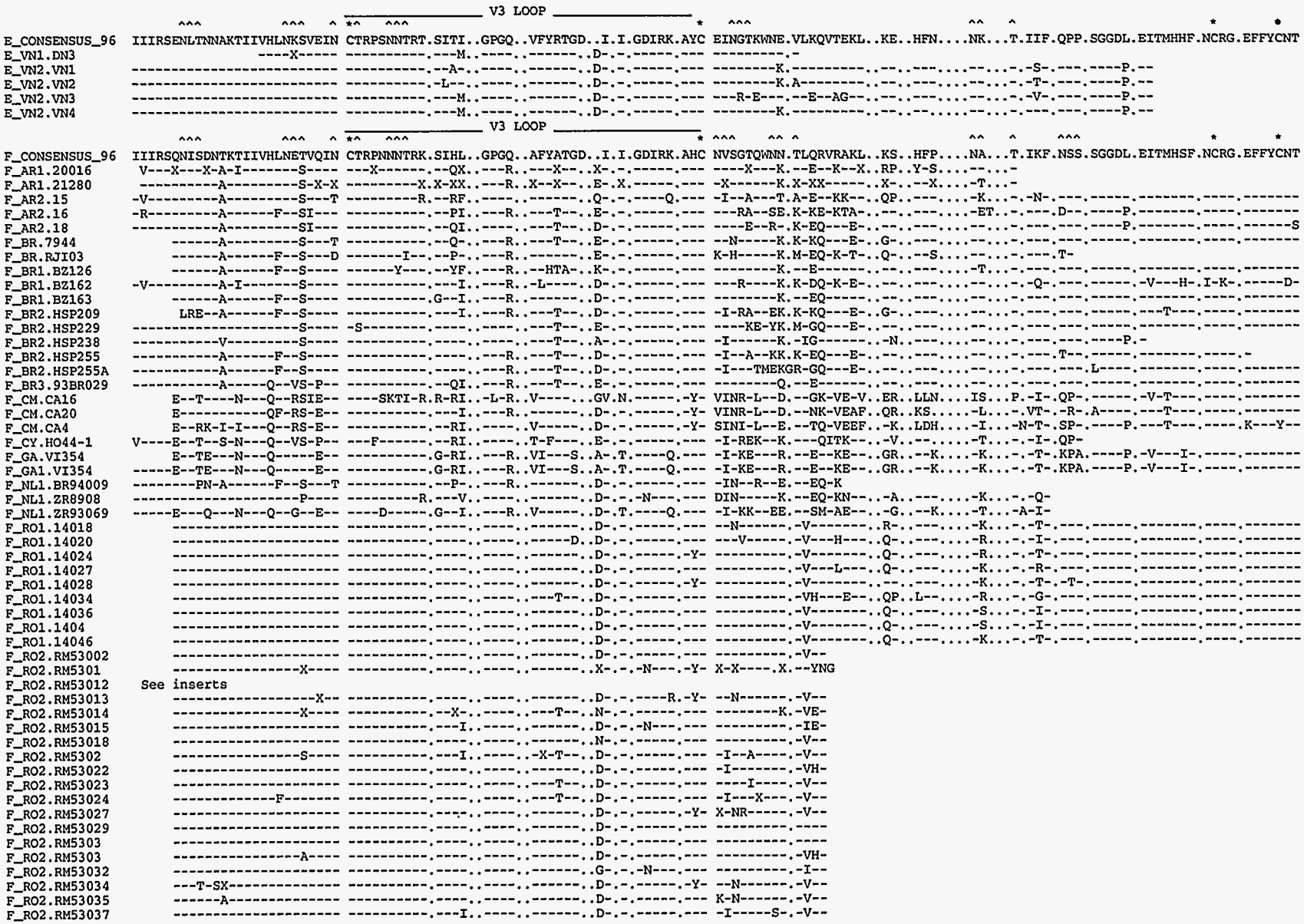
V3 LOOP
C_CONSENSUS_96 IIIRSENLTNNVKTIIVHLNESVEIV CTRPNNNTRK.SIRI..GPGQ..TFYATGD..I.I.GDIRQ.AHC NISKDKWNE.TLQRVGKKL..AE..HFP....NK...T.IKF.APS.SGGDL.EITTHSF.NCRG.EFFYCNT
C_IN4.CMCH11
C_IN4.CMCH12
C_IN4.CMCH13
C_IN4.CMCH15
C_IN4.CMCH19
C_IN4.CMCH2
C_IN4.CMCH20
C_IN4.CMCH22
C_IN4.CMCH23
C_IN4.CMCH26
C_IN4.CMCH27
C_IN4.CMCH29
C_IN4.CMCH3
C_IN4.CMCH32
C_IN4.CMCH33
C_IN4.CMCH36
C_IN4.CMCH37
C_IN4.CMCH4
C_IN4.CMCH5
C_IN4.CMCH6
C_IN4.CMCH7
C_IN4.CMCH8
C_KE.NA113
C_MW.SM750
C_MW1.12199
C_MW1.12203
C_MW1.12205
C_MW1.12209
C_MW1.12213
C_MW1.12215
C_MW1.12225
C_MW1.12229
C_MW1.12233
C_MW1.1227
C_MW1.6508
C_MW1.6512
C_MW1.6518
C_MW2.D3MA959
C_MW2.D3MA960
C_MY.9214082
C_MY.9214083
C_NL1.NL94024
C_NL1.UN93050
C_NO1.V3N13
C_NO1.V3N14
C_NO1.V3N17
C_NO1.V3N19
C_RU.RUS13A
C_RU.RUS1A
C_RU.RUS20A
C_RU.RUS2A
C_RU.YAN4
C_RW1.134cons
C_RW1.566cons
C_SN.SE364
C_SO.1574
C_SO.SM145
C_TW1.252

```
                                                  V3 LOOP
                 ^^^           ^^^   ^ **    ^^^                          * ^^^    ^^ ^                  ^^^       ^ ^^                          *       *
C_CONSENSUS_96   IIIRSENLTNNVKTIIVHLNESVEIV CTRPNNNTRK.SIRI..GPGQ..TFYATGD..I.I.GDIRQ.AHC NISKDKWNE.TLQRVGKKL..AE..HFP....NK...T.IKF.APS.SGGDL.EITTHSF.NCRG.EFFYCNT
C_UG.45          ------D-N--A-I----------N ---------E.-V--..-----..A----K-..-.-.-----.--- HNNGT---A.---K-RE--..KQ..Y--....DR....-.-E-.NN-.A-----.-------.-----.-------
C_UG.UG268       ----------A-I---Q--K----N -A--------E.-----..-----..-------..-.-.-----.-Y- ---RNE--I.---W-RE--..KR..---.....--....-.-N-.TQP.-----.-------.-----.------
C_UK1.00513      ---S-----D-A-I---------- ----------.-V--..-----..A----N-..-.-.-----.--- --..TG-DK.-M-Q
C_ZA.NOF         ----------I---V----K--D-- ----------.R--V..-----..-V---NA..-.-.-----.--- ----KL---.--EQ-K---..--..-Y-....--....-.-A-.NS-.-----
C_ZA.ZA514       -------I---------------- -I--G-----.-V--..-----..--F---E..-.-.-K--E.--- ---E-Q--K.--H--SE--..E-..---.....--....-.---.G-P.T----.-------.-----.-------
C_ZA.ZA517       ----------A------------- ----GS----.----..-----..A-F----..-.-.-----.--- ---EN---K.---K-RR--..E-..Y--....--....-.---.ER-.----I.-VV----.--G-.-------
C_ZA2.BooyD      ----------I----A-----IK-- ----------.----..-----..A-H--N-..-.-.-----.--- ----V---T.--EK-K----..K-..---.....--....-.-E-.K-H.------.-V-----.----.-------
C_ZA2.Dlu        ---------D---------------E ----------.----..-----..-----NG..-.-.-----.-Y- ---TRA--K.----------..K-..N--....-S...-.---.E-H.---PP.-------.-----.-------
C_ZA2.GOM        --V------D---------------- ----S-----.----..-----..A----N-..-.-.----E.--- ---T----K.--EQ------..E-..---.....--....-.-T-.K-H.-----.-------.----.-----Y-
C_ZM1.ZAM18      ----------A------------- ----S-----.----..-----..A-----G..-.-.-N---.--- ----EN--K.---K-----..--..---.....--....-.---.DQH.-----.-------.----.------
C_ZM1.ZAM20      ----------AQ----Q--K-I--- -A--G-----.----..-----..--F---A..-.-.-----.--- ---RSE---.--PN-SR--..-G..---....-R....-.-I-.E--.-----.-------.----.-----
C_ZW.2647         --------D----------KP-T-- ----------.----..----..-------..-.-.-N---.--- --N-TE--K.--E--KE--
                                                  V3 LOOP
                 ^^^           ^^^   ^ **    ^^^                          * ^^^    ^^ ^                  ^^^                           *       *
D_CONSENSUS_96   IIIRSENLTNNAKIIIVQLNESVTIN CTRPYNNTRQ.RTHI..GPGQ..ALYTT....I.I.GDIRQ.AHC NISGAEWNK.TLQQVAKKL..GD..LL.....NKT..T.IIF.KPS.SGGDP.EITTHSF.NCGG.EFFYCNT
D_BI.BU009con    ------------------------- -----K-I-R.G---..-L-R..-Y---.G..-.E.-KKG-.P-- K-N--G---.-----I---..R-..-F.....--....-.---.--P.----L.-------.-----.-------
D_CF.402019      --V------D---T-----K---N-S ----N-YK--.G-P-..-L--..------.R..V.-.----K.--- ---RK---N.---K------..RE..-F.....--S..K.-N-.Q-H.----L.-V-----.--Q-.------
D_CI.CI13        ------------------------- ----N-----.G---..-----..-----.G..V.-.----K.--- ----KA---.-----I-I---..RN..--.....---..-.---.NS-.---G-.--I----.-----.------
D_GA1.G109       ------D------N------V--S-- ----G-K---.K-P-..-L--..----S.K..M.-.V-R-K.--- -V-ATD--N.---------..--..--.....---..-.-N-.-S-.-----.-------.----.-------
D_GA1.G141       ---------D-T-N------D-IV-- -I--N-----.S---..-----..--F---R..V.-.-----.-Y- E-N-TK--N.------R---..-N..--.....---..-.---.-----.-V-----.--R-.-------
D_KE.NA116       ------D-------------K----D -I--------.G---..---R..-YF---.E..-.T.-----.-Y- K-N-TA---.--E------..RG..-V.....--...-A---.---
D_KE1.KEN966     ------------------------- ----------.S-R-..-----..T-F---.K..-.-.-----.--- -----G---.---R-----..--..--
D_KE1.KEN97      ----------T---------P--- ----------.G---..---R..-Y---.N..-.-.-----.--- ---R-----.-------R-..--..--
D_KE1.KEN986     ----------------------P-T ----------.G---..-----..T-F---.K..-.-.-----.--- ---E-----.---L--R---..-G..--
D_NL.A1          -----Q-ISD-T-T---H---T-Q-- ----N----R.SI-V..-----..-F-A-GD..-.-.-N--K.--- D-N-TQ---.--E--KN---..KT..HFP....--....-.-Q-.NQ
D_NL2.AO93021    -------I-H-T-N---H-----V-- ----------.S---..-----..--F---NR..-.-.-----.--- S----D--R.---R--T---..-N..--.....-T-..-.-K-
D_NL2.UG93071    -V----------T------K-IE-T -A--S----K.GI--..-----..-YF-A.E..-.T.-----.--- ----TN-SV.A---------..--..--.....---..-K-T-
D_NL2.UG94015    -M---------------------I-- ----N-----.SI--..-----..------.N..-.-.-----.--- ---EKA---.----------..--..--.....-R-..-.---
D_NL2.ZR8911     ---------D-T-N-----KK-IS-K -I--N-----.S---..-----..------.Q..V.-.----K.-Y- ---KT--DN.-------E-..-K..--N....TRI..-.-K-
D_NL2.ZR9011     ---------D-T-N------K-IE-E ----N-----.S-P-..-----..------.K..V.-.------.-Y- T-NRT--DS.---K------..-N..--.....---..-.---
D_NL2.ZR9291     ----------T---------E-T -R---------.G---..-----..------.E
D_NL2.ZR94022 See inserts
D_RU.RUS14A      ------------------------- -------I--.--P-..-L--..------.R..R.-.E---R.--- ---K-A-K-.--E-------.KV..-F.....-R-..-.-D-.-
D_SN.SE365       ----------T----F----M-- -------K---.--P-..-L---..V-H--.R..V.K.------.--- ---KER---.-----VR---.K-..-F.....--....-.---.E--.------.--A-----.--R-.-------
D_TZ1.TAN        -------F----------I---- -----SRV--.GA---..---R..-Y-A-.N..-.F.------.-R- -----A---.--ER--N---.R-..--N....-QP..-.---.--
D_TZ1.TAN1       --V-------------------S-- ----S------.S-R-..-----..----N.K..-.-.-N---.--- -V-KIK---.------N---.RE..--.....---..-.-N-.--
D_TZ1.TAN12      ----------T---------Q-T -A---------.S-R-..-----..--F-S.K..-.-.-N---.--- --TR-A---.----------..-G..--.....-Q-..-.-V-.--
D_TZ1.TAN13      --------A----N---------- -----E-M--..--P-..-L--..--V-S.R..-.K.-R--P.-Y- ----ES-HR.----------..--..-F.....---..-.-L-.--
D_TZ1.TAN2       ----------------------P-- ----------.G---..---R..-Y---.R..-.L.-N---.--- S--VIK---.----------..R-..--.....-I...-K---.--
D_TZ1.TAN3       --------------------K----- --------I-.G---..---R..-Y---.D..-SVI------.--- --------E.----------..--..--.....-Q-..-.---.--
D_TZ1.TAN4       ------------------------- --------I-.K-S-..-R--..------.E..T.R.---K-.-F- K---IA---.----------..R-..-F.....---..-.---.--
D_TZ1.TAN5       ----------N-------------- -------I--.--P-..-S--..------.R..R.-.------.-Y- E--E-G---.----------..R-..--.....---..-.----.-S
D_TZ1.TAN6       ----------------------P-- ----------.G---..---R..-Y---.R..-.V.-N---.--- S--VIK---.----------..R-..--.....-I...-K---.--
D_TZ1.TAN7       ------------------------- ----------.S--F..---R..-Y---.D..-.-.-------.--- ----IG---.----------..R-..--.....---..-.---.--
D_TZ2.005        -------F----------------A-- -----R----.-----..-S--..------.R..-.TG----R.--- T---E----.--H-------..R-..--.....---..-.---.---.-----.-------.---
D_TZ2.012        ------D------V-----------K -S---Y-Q-R.-----..-T--..------.R..-.R.------.--- ---E-G---.------E---..R-..-F.....---..-E---.R--.-----.-------.---
D_TZ2.023     See inserts
D_TZ2.030        ------------------------- -----K--I-.----..-S--..--F---.R..G.-.--P--.-Y- ---K-----.---R------..--..-F.....---..-.---.---.-----.-------.---
D_TZ2.053     See inserts
D_TZ2.064     See inserts
D_TZ2.080     See inserts
D_TZ2.112        ------------------------- ----FTR---.--S-..-Q--..-F-RATR..-.-.------.-Y- ---------.----------..--..-F.....-Q-..-.---.--H.-----.-------.-----.S
D_TZ3.1585            -------L-----KAP-M-- -S--N----K.SIR-..-----..TF-A-GG..-.-.-----E.--- ---RTD--N.---E------..RV..--.....---..-.---.-----.-------.--Q-.-------
D_TZ3.1604            ------------------- ----N-----.S---..----..------.K..-.-.-----.--- K--R-----.----------..R-..--.....---..-.---.N--.-----.-------.----.------A
D_TZ3.1627    See inserts
D_TZ3.4622    See inserts
D_UG.U44342   See inserts
D_UG.UG23           --------T-------T---- -----E-V-H.--P-..-L--..--I-N.R..-.K.AK-G-.-Y- ---KT--D-.---R--T--..R-..--N....HT...-.-N-.---.----Q
D_UG1.W2UG00  See inserts
D_UG1.W2UG005        ------------------- ---------K.GI--..---R..-Y---.N..-.V.-N---.--- -V-K-K--N.----------..R-..--.....---..-.-T-.R--.-
D_UG1.W2UG02         ----------H-----P-- -----DKVSY.--P-..-V-R..-S---.R..-.K.------.--- ----EK---.------A--..R-..--.....-Q...-A---.---.-
```

```
                                                                  V3 LOOP
                    ^^^          ^^^   ^ *^    ^^^                          * ^^^    ^^ ^                    ^^^                                  *       *
D_CONSENSUS_96  IIIRSENLTNNAKIIIVQLNESVTIN CTRPYNNTRQ.RTHI..GPGQ..ALYTT....I.I.GDIRQ.AHC NISGAEWNK.TLQQVAKKL..GD..LL.....NKT..T.IIF.KPS.SGGDP.EITTHSF.NCGG.EFFYCNT
D_UG1.W2UG024          ----------------V-- -------I--.--P-..-L--..-----.R..R.-.E---R.--- ---E-A-K-.--E------..KV..-F.....-R-..-.-D-.---.-
D_UG1.W2UG035          -------------KK----K -I-------E.G---..---R..--F--.D..-.-.-----.--- ---RT---E.---------..--..-F.....--...E.---.---.-
D_UG1.W2UG038 See inserts
D_UG1.W2UG040 See inserts
D_UG1.W2UG046          -------T---------P-- -S---E-K-R.--P-..-L--..-Y---KL..K...-Y--P.--- ---------.---------..--..-F.....-Q-..-.---.Q-H.-
D_UG1.W2UG053 See inserts
D_UG1.W2UG059          ----------------P-- -----R-I--.----..-T--..-Y-.-KG..-.K.-VAG-.P-- I-N--K---.---------..--..-F.....---..-.---.-
D_UG1.W2UG065 See inserts
D_UG1.W2UG070 See inserts
D_UG2.1665     See inserts
D_UG2.1685     See inserts
D_UG2.2999        -------IS----T-------T---- -----H-KI-.----..-T--..--H--.R..-.T.-Y-G-.--- K--I-K-SN.------T--..R-..-F.....-R-..-.-N-.---.-----
D_UG2.3           ----------V-T----F--T-K-- -----Y-I--.--P-..-L--..-----KG..RGTTKV-G-.--- ---------.---R--I--..-Y..--.....---..-.---.NS-.A----
D_UG2.4132        ----------T----------- ------K-S-.--S-..-Q-R..-----.K..P.T.-Y---.-Y- T-N--R---.AW
D_UG2.4133        ----------N----------- -S---KS--I.----..-S--..-Y-R-.N..-.Q.-----.-Y- -----A---.--K------..--..--.....--...-.T--
D_UG2.462         ---------------------- -----RAM-R.--S-..-Q--..-Y---NITG-.G.-N---.-Y- -----G---.I
D_UG2.5055        ----------V-T-------T-K-- -----S-K--.S-P-..-L--..-----RG.R.....----K.--- -V--------.---R--I--..--..--.....---..-.---.NT-.-----
D_UG2.5059        ----------V-T----F--T-K-- -A---EKK-R.T-P-..-L--..--I-SRNPEK.....-G-.--- ---------.---R--I--..--..--.....-Q-..-.---.-S-.-----
D_UG2.653      See inserts
D_UG2.G        See inserts
D_UG2.G2       See inserts
D_UG3.109         ----------N----------T -M-----K--.SV--..---R..-F---...N-.-.-----.--- ---KVG-D-.-----RG--..K-..-F.....-T...-.-N-.---
D_UG3.110         ----------------------P-- ----------.G---..---R..-YW---...N-.-.-----.--- ---R-----.---------..--..--.....---..-.---.---.----
D_UG3.114         ------------------------- ----------.GI--..---R..-Y-.-D..Q-.T.-----.--- ---K-----.---------..RE..--.....---..-.---.---.---
D_UG3.12       V----------T----------A-- ----N-----.S---..----...-----...K-.-.----R.--- E--E-G---.--------R---..-H..--.....-T...-.---.---
D_UG3.120         ---------D----------A-- ----S----K.SI--..----..--F-I...D-.-.-N---.--- ---R-K---.---------..--..-F.....-T...-N---.--
D_UG3.122         ----------T---R---T-K-- ----A-----.SV-L..----..-----...RV.-.-----.-Y- S--E----R.------V--..--..--.....-Q-..-.-N-.---.----
D_UG3.7        -VV----------H--K----- ----N-----.SI--..----..-----...NV.-.-----.--- ---R-----.-----TE---.KY..-F.....--...-I-N-.E
D_UG3.70       See inserts
D_UG3.74          -------F----------K----T -----T-K--.G--M..---R..----I...D-.T.-----.-Y- ---K-K-YN.------N--..--..--.....---..-.-N-.E--.----
D_UG3.79          ----------T----------- -----------.S---..----..-----...N-.-.-----.--- -V-K-G-G-.---------..K-..-FK....AD...N.-T-.---.----
D_UG3.83          -------I----RT-----K---K-T -V---S-Q-R.--P-..-L--..-----RMDNM.K..N-K-.-Y- ---VED---.---------..-G..--N....LT...-.---.---.----
D_UG4.C971-11     -------I----------------- -----QT--R.--P-..-L---..-V--S...K-.A.-YAGP.-Y- --N-IA---.---------..--..--.....-R-..-.-L-.Q-P.-----.-----T-.----.--S----
D_UG4.WHO1523     ----------T---H----IQ-- -----H--G-.----..-T--..-----...R-.-.---SK.--- ---E-G--T.---------..-Y..--.....---..-.---.---.-----.-------.----.-------
D_UG5.UG266    See inserts
D_UG5.UG269       ---------------------A----T -----TK--H.-.A....Q-R..-WW---...G-.T.-----.-Y- ---K-K-QE.---------..--..--.....---...-Q---.E--.-----.-V-K---.----.-----D.
D_UG5.UG274       ------------------------D -A---E-I--.--P-..-T--..-----QGRKK.-.-K-G-.--- -V-K-G-E-.-------N
D_UG6.980      See inserts
D_UG6.9802           ---------------------- -----QTI-H.----..-S--..----P...R-.T.-Y---.--- -V--------.---------..--
D_UG6.9803           ---------------------- -----K-I-H.--P-..-S--..-----...R-.A.V-----.--- ---K-----.---------..--
D_UG7.42       V------I---V-T---L---T-Q-- -----SKI--.--ST..-Q--..-----KGRG-.-.-...-.-Y- ---------.--HS.....SN..I-RRP.PSQ...-K-N-LPHP.LE-T-.--------.I---.---S--
D_UG7.44          ----------T-----K-P-N-T -I-------K.S---..----..-YF-S...E-.-.-----.--- ---K-Q--N.--T---N--..-G..-S.....ST...-M---.S--.-----GKVHQ--V.-----.G----D
D_UG7.963      See inserts
D_UG8.94UG010                    ---- ----------.S--M..---R..------...RT.-.----K.--- K--R----T.---RL-I--..--..--K....QT...-.-S-.NT-.------.--------.----.-----
D_UG8.94UG013                    ---T -A--------.G---..---R..-Y---...K-.-.-----.--- -----G--N.------R---.K-..--.....---..-.---FP--.--AH-.V------.A---.---C
D_UK1.CPHL4    -------I-D-V-N---------L-T -A--S-------.S---..----..-----G..Q-.-.-----.--- -V-KKD-SN.---R--I---.KN..--.....---..-.----.NT-.------.--------.----.--------
D_US.AMK       -------I-----T---------S-- -----SKE-L.K-S-..-Q--..-----RVKV..T.-----.--- -V-EIK---.------A---.KG..--N....ISS..-.-H-F---.------.--------.----.------
D_ZA.ZA500     -------------N------V--E-- ----------.N-K-..-----..TFFARK..M-.-.-----.--- ---AEK---.------V---..-N..--.....---..-.-N-.R--.------.--------.---K.-----Y-
D_ZA.ZA501     -------------N----F-A--E-- -----KYIS-.--S-..-Q---..V-H-SK..K-.-.-----.-Y- ----EK---.------IE---..-N..--.....---..-.----.Q-P.----S.--------.----.--------
D_ZA.ZA505     -------------N----F-A--E-- -----RYI--.K-S-..-Q---.T-H-SK..R-.-.-----.--- ----RK-H-.------T---..RN..--.....---..-.----.R-P.P----.--------.----.--------
D_ZA.ZA506     -------------N----F-A--E-- -----QYA--.K-S-..-Q---.T---SK..K-.-.-----.-Y- ---EEK---.------I---..-K..P-.....---..-.----.R-T.----S.--------.----.--------
D_ZA.ZA507     -------------N----F-A--E-- -----EIRI-.K-S-..-Q---..-N-NK..R-.-.-----.-N- T---EK---.------I---..-N..--.....---..-.-A-.R-T.------.------V.----.--------
D_ZR.ELI       V------------N--AH-----K-T -A---Q----.--P-..-L---.S----RSRS-.-.-...-.--- ---R-Q-S-.------R---..-T..--.....--...-I-K-.--
D_ZR.JY        See inserts
D_ZR.MAL          -M-------D-T-N-------T---- ----G----R.GI-F..-----..------...G-.V.----R.-Y- T-NET--D-.------V---..-S..--.....---...-K---.NS
D_ZR.NDK          ----------V-T------A-IV-- -----KY---.--S-..-LR-..S---ITGKKKKT.-Y-G-.--- K--R------.A-----T---..-N..--.....---..-.-T-.--
D_ZR.Z2Z6         ----------------------A-- -----R-I--.--S-..-L---..-----KTRS-.-.-...-.-Y- ---KN----.------I---..-N..--.....---..-.---.--
D_ZR1.6555           -K-I---VNN---------A-- -I--N-----.S--L..----..------...KV.-.-----.-Y- ----T-----.A-----I---..-N..-F.....---..-.-N-.DH-.------.------
D_ZR1.6565.sh     ---------D-L-N---------I-- ----N----K.SI--..----..-I---...DV.-.-----.-Y- E-N-PK---.----
D_ZR1.6651.sh        ---------T------AT-A-- ----N----K.GI--..----..----SG..D-.V.-----.---
D_ZR1.6659.sh     -------D---T------------ ----N--V--.--P-..-----..-F---G
```


V3 LOOP
E_CONSENSUS_96 IIIRSENLTNNAKTIIVHLNKSVEIN CTRPSNNTRT.SITI..GPGQ..VFYRTGD..I.I.GDIRK.AYC EINGTKWNE.VLKQVTEKL..KE..HFN....NK...T.IIF.QPP.SGGDL.EITMHHF.NCRG.EFFYCNT
E_CF.4081 -------I-D---I------E----- ----N----K.-VR-..-----..T--A--..D-.-.----Q.-H- NVSR----Q.T-Q-IATQ-..RK..Y-G....-TT..-.-N-.SKS.------.---T-T-.--G-.-------
E_CF.90CR402 ------D--D----------I--- ----FKKV-I.-AR-..---R..--HT--..N-.N.-----.--- ---K---K-.T-----R--..R-..-L-....GTM..-.-S-.R-S.----P.-------.----.-------
E_CF1.1697 QNNQIX---------- ----FK-M--.-AR-..-----..---K--..S-.T.-----.--- ---------.T-----K--..R-..--K....--...-.---.-S
E_CF1.4002 ---------D----------I--- ----FKKV-I.-AR-..---R..--HT--..N-.N.-----.--- ---E---K-.T-----R--..R-..-L-G...TT...-.-S-.TTT
E_CF1.4013 ----------E--T-- ----FKRV--.-VR-..---R..--HK--..A-.N.-----.--- K-D-------.T--R--I--..R-..--.....--...-.-LY.R-
E_CF1.4017 ----------A----- ------KI--.-VR-..-----..---K--..A-.M.-----.-F- ---------.T--E-VK--..RK..--.....--...-.-H-.R-
E_CF1.403 ---------- ----------.-VR-..-----..---K--..D-.-.----R.--- ---------.T-----K--..R-..Y-.....--...-.---.K-
E_CF1.4039 VV-----I---------Q-VTP-Q-- -------I--.-VR-..-----..---K--..S-.T.-----.-F- --KEAQ--K.T--E-IG--..R-..--.....--...-.---.-R
E_CF1.4069 --------AD----------E----- ----FKK--I.-AR-..---R..--HK--..A-.L.-----.-F- ---K-R--D.T-NK-----.--..--....K-...-.-L-.--
E_CF1.4084 ..N------D----------P---- Y---------.-VR-..-----..---K--..E-.-.-----.-F- K--E-----.T--E--K--..R-..--K....--...-.---.--
E_CF2.4071TG2 ---------- ----FKKM--.-VR-..---R..---K--..S-.T.-----.--- ---------.T-Q-IIR--..E-..---....--...-.-Q-.K--Y-----.-------.----.-------
E_CF2.ELO ----------R---------- ----YK----.-AR-..-----..---K--..S-.T.-----.--- -----A-H-.T---MVG--..RK..--.....--...-.---.--S.----P.-----N-.-----.-------
E_CF2.MBA --------A----------P---- ----------.-VR-..-----..---G--..E-.-.----R.-F- ---------.T--------..R-..--.....--...-.-V-.---.-----.-----S-.-----.-------
E_CM.CA10 ---------- -----TTI--.-SR-..-----..A--KIE..G-.S.-N--A.--- ---------.TF----K--..R-..Y-....--...-.-K-.--S.P---P.-----S-.-----.-------
E_JP1.JNIH2T ----D----------I--- ----------.-----..-----..-------..D-.-.-----.--- ---------.T--------
E_JP1.JNIH3J ----D---------- ----------.-----..-----..-S-----..D-.-.-N---.--- -----E---.T----A---
E_JP1.JNIH4T ----D---------- ----------.-----..-----..-------..D-.-.-----.--- --------K.---------
E_MY.1786 ---------- ----F-----.-----..-----..-------..D-.T.-----.--- --------K.A---------..-
E_MY.9214103 ---------- ----------.-----..-----..-------..D-.T.-N---.--- ---------.---------..--.--.....--...-.---.---.-----.-------.----.-------
E_NL.TH94037 ------D---------- ----------.--H-..-----..-------..D-.T.----Q.-S- V----W--V.--QEAA.E-..F-..--F....--...-.-F-
E_TH.93TH253 ----------HE----- ----FY-K--.RTS-..-Q-R..-L-----..D-.T.-N-G-.P-- --------K.--N------..--..---.....-R...N.-S-.---.-----.-------.I---.-------
E_TH.CM240 ------D---------- ----------.-----..---R..-------..D-.-.-N---.--- --------K.---------..--..--.....--...-.---.---.-----.-------.----.-------
E_TH.JP23A ---D----------T----- -----THK-K.RM-L..---H..---S--..D-.T.-----.--- --------K.A---------..--..--I....--...-.-V-.---.-
E_TH.N764 ----------K ----------.-----..-----..-------..D-.-.-N---.--- ----------.T----AK--..N-..--.....K-...-.---.H--.-----.VV-----.-----.-------
E_TH.T8178 ---------- ----------.--S-..-----..-------..D-.-.-----.--- ----------.A-T----N-..--..--H....--...-.-L-.---.-----.---T---.-----.-------
E_TH1.8639 T---------- ----------.R---..-----..-------..D-.-.-----.--- ----------.A---------..--..--...--...-.---.---.-----.-L-----.-----.-------
E_TH1.8657 ---------- ----------.-----..-----..I------..D-.-.-----.--- ----------.A---------..--..--...--...-.---.---.-----.-------.-----.-------
E_TH1.8659 ---------- ----------.-----..-----..-------..D-.-.-----.--- ----------.---------..--..--...--...-.---.---.-----.-------.-----.-------
E_TH1.8663 ---------- ----------.-----..-----..-------..D-.-.-----.--- --------K.---------..--..--...--...-.---.---.-----.-------.-----.-------
E_TH1.867 ---------- ----------.-----..-----..-------..D-.-.-N---.--- ----------.A----AG--..--..--.....--...-.---.H--.-----.VV------.--S-.-------
E_TH1.8673 ---------- ----------.-----..-----..-------..D-.T.-N---.--- ----------.A-R-------..--..--...--...-.---.---.-----.-------.-----.-------
E_TH1.8683 ------E---------- ----------.--P-..-----..-------..D-.-.-N---.--- ----------.T---------..--..--...--...-.---.---.-----.-------.-----.-------
E_TH1.TA1067 ---------- ----------.---M..-----..-------..D-.-.----R.--- ----K---K.A---------..--..--I....--...K.---.---.-----
E_TH1.TA7792 ---------- ---------P.-----..-----..-------..D-.-.-----.--- ---------K.A---------..--..--...--...-.--C.---.---P.-
E_TH1.TA7794 ---------- ----------.-----..-----..---K--..D-.-.-N---.--- ----------.A---------..--..--...--...-.---.---.-----.--------.-----
E_TH1.TA8173 ---------- ----------.--P-..-----..-------..D-.-.-----.--- ----------.A---------..--..--...--...-.---.---.-----.---T---
E_TH1.TA8176 ---------- ----------.-----..-----..-------..D-.-.-----.--- -V--------.A---------..--..--...--...-.---.---.-----.-------
E_TH10.92TH001 ---------- ----------.--N-..-----..-------..D-.-.-----.--- ----------.--R------..--..--I....-R...-.---.---.-----.-------.-----.-------
E_TH10.92TH011 ----------I---------- ----------.-----..-----..-------..D-.-.-N---.--- --------Y-.------K--..--..--T....--...-.---.---.-----.---T---.-----.-------
E_TH2.TN235 ---------- ----------.--P-..-----..A------..D-.-.-----.--- ----------.--T-------..--..--...--...-.---.---.-----.-------.--X-.-------
E_TH2.TN238 ---------- ----------.-----..-----..-------..D-.-.--T--.--- ---S-R---.A-----G--..--..--...--...-.---.---.-----.-------.-----.-------
E_TH2.TN239 ---------- ----------.-----..-----..-------..D-.-.-N---.--- ----------.A---------..--..--...--...-.---.---.-----.-------.-----.-------
E_TH2.TN24 ---------- ----------.-----..-----..-------..D-.-.-----.--- ----------.------GE--..--..--...--...-.---.---.-----.-------.-----.-------
E_TH2.TN240 ------D---------- ----------.-----..-----..-------..D-.-.-N---.--- --------K.---------..--..--...--...-.---.---.-----.-------.-----.-------
E_TH2.TN242 ---------- ----------.-----..-----..-------..D-.-.---X-.--- ---R------.A-----K--..--..--...--...-.-L-.---.-----.-------.-----.-------
E_TH2.TN243 ---------D----------G-- --------XP.---X..-----..-------..D-.-.----X.--- --------R.---------..--..--...--...-.---.---.-----.-------.-----.-------
E_TH2.TN244 ----------V-- ----------.-----..-----..-------..D-.-.-----.--- -----E--K.A---------..--..--...--...P.---.---.-----.-------.-----.-------
E_TH3.W2TH00 ---------- ----------.--N-..-----..-------..D-.-.-----.--- ----------.--R------..--..--I....-R...-.---.---.-
E_TH3.W2TH003 ---------- ----------.---V..-----..-------..D-.T.-----.--- -V--------.T-----KN-..--..--...--...-.---.---.-
E_TH3.W2TH005 ---------- ----------.--P-..-----..-------..D-.-.-----.--- --------K.---------..--..--K....--...-.---.---.-
E_TH3.W2TH006 ---------- ----------.-----..-----..-------..D-.-.----Q.--- ----E-----.T-----K--..--..--...--...-.---.X--.-
E_TH3.W2TH007 -----V---------- ----------.--N-..-----..-------..D-.-.-----.--- --------K.---------..--..--...--...-.---.-M-.---.-
E_TH3.W2TH009 ---------- ----------.-----..-----..-------..D-.-.-----.--- ---R------.A-----K--..E-..--...--...-.---.---.-
E_TH3.W2TH01 -------I---------- ----------.-----..-----..-------..D-.-.-N---.--- -------Y-.------K--..--..--T....--...-.---.---.-
E_TH3.W2TH015 ---------- ----------.G---..-----..-------..D-.-.-----.--- --------K.-----X----..--..--...--...-.---.-
E_TH3.W2TH018 ---------- ----------.-----..-----..I------..D-.-.-----.--- ---------.A---------..--..--K....--...-.---.-
E_TH3.W2TH019 ----------K-- ----------.-----..-----..-------..D-.-.-----.-H- -----E---.A-----K--..--..--...--...-.---.---.-
E_TH3.W2TH02 ---------- ----------.-L---..-----..-------..D-.-.-----.--- --------K.A---------..--..--...--...-.---.---.-
E_TH3.W2TH020 ---------- ----------.-L---..-----..-------..D-.-.-----.--- ---------.A---------..--..--...--...-.---.---.-
E_TH3.W2TH022 ---------- ----------.-----..-----..-------..D-.-.----R.--- --------K.A---------..--..--...--...-.---.---.-

V3 LOOP
E_CONSENSUS_96 IIIRSENLTNNAKTIIVHLNKSVEIN CTRPSNNTRT.SITI..GPGQ..VFYRTGD..I.I.GDIRK.AYC EINGTKWNE.VLKQVTEKL..KE..HFN....NK...T.IIF.QPP.SGGDL.EITMHHF.NCRG.EFFYCNT
E_TH3.W2TH023
E_TH3.W2TH024
E_TH4.D3TH966
E_TH4.D3TH975
E_TH4.D3TH976
E_TH5.CMU0
E_TH5.CMU02
E_TH5.CMU03
E_TH5.CMU04
E_TH5.CMU05
E_TH5.CMU06
E_TH5.CMU07
E_TH5.CMU08
E_TH5.CMU10
E_TH5.E1058
E_TH5.KH00
E_TH5.KH003
E_TH5.KH004
E_TH5.KH005
E_TH5.KH007
E_TH5.KH008
E_TH5.KH009
E_TH5.KH01
E_TH5.KH012
E_TH5.KH013
E_TH5.KH014
E_TH5.KH015
E_TH6.0182
E_TH6.1012
E_TH6.194
E_TH7.1018
E_TH7.1110
E_TH8.0103
E_TH8.0289
E_TH9.100
E_TH9.106
E_TH9.111
E_TH9.115
E_TH9.119
E_TH9.126
E_TH9.135
E_TH9.179
E_TH9.81
E_TH9.83
E_TH9.84
E_TH9.89
E_TH9.97
E_TH9.98
E_TW1.286
E_TW1.346
E_TW1.396
E_UK1.11643
E_US.POC30506
E_UY.UR2
E_UY.UR4
E_UY.UR6
E_UY.UR7
E_VN1.BX6
E_VN1.CT5


V3 LOOP
E_CONSENSUS_96
E_VN1.DN3
E_VN2.VN1
E_VN2.VN2
E_VN2.VN3
E_VN2.VN4
V3 LOOP
F_CONSENSUS_96
F_AR1.20016
F_AR1.21280
F_AR2.15
F_AR2.16
F_AR2.18
F_BR.7944
F_BR.RJI03
F_BR1.BZ126
F_BR1.BZ162
F_BR1.BZ163
F_BR2.HSP209
F_BR2.HSP229
F_BR2.HSP238
F_BR2.HSP255
F_BR2.HSP255A
F_BR3.93BR029
F_CM.CA16
F_CM.CA20
F_CM.CA4
F_CY.HO44-1
F_GA.VI354
F_GA1.VI354
F_NL1.BR94009
F_NL1.ZR8908
F_NL1.ZR93069
F_RO1.14018
F_RO1.14020
F_RO1.14024
F_RO1.14027
F_RO1.14028
F_RO1.14034
F_RO1.14036
F_RO1.1404
F_RO1.14046
F_RO2.RM53002
F_RO2.RM5301
F_RO2.RM53012
See inserts
F_RO2.RM53013
F_RO2.RM53014
F_RO2.RM53015
F_RO2.RM53018
F_RO2.RM5302
F_RO2.RM53022
F_RO2.RM53023
F_RO2.RM53024
F_RO2.RM53027
F_RO2.RM53029
F_RO2.RM5303
F_RO2.RM5303
F_RO2.RM53032
F_RO2.RM53034
F_RO2.RM53035
F_RO2.RM53037


V3 LOOP
F_CONSENSUS_96
F_RO2.RM53040
F_RO2.RM53043
F_RO2.RM5306
F_RO2.RM5307
F_RO2.RM5308
F_TW1.334-1.sh
G_CONSENSUS_96
G_BJ1.259
G_BJ1.43
G_CF.4067
G_GA.LBV21-7
G_GA.VI525
G_GA1.G98
G_GA1.VI526
G_GB.K
G_GB1.22
G_KP.Kr121
G_NG1.G3
G_NG1.G9
G_NG1.JP882
G_NG1.JV832
G_NL.127C
See inserts
G_NL1.GH94012
G_NL1.LR94018
G_NL1.UM92101
G_NL1.ZR9119
G_RU.131con
G_RU.RUS12A
G_RU1.BUK3a
G_TW1.267
H_CONSENSUS_96
H_CM.CA13
See inserts
H_ZR.VI557
I_CONSENSUS_96
I_CY-HOcon
J_CONSENSUS_96
J_GM1.GM4
J_GM1.GM5
J_GM1.GM7
J_SE1.7022
J_SE1.7887

```
                                                  V3 LOOP
                  ^^^          ^^^   ^  *^   ^^^                  *   ^^^                          ^^   ^ ^^^     ^^^           *       *
O_CONSENSUS_96    IRIMGKNISDSGKNIIVTLNSTINMT CERPGNQTVQ.EIKI..GPMA..WYSMGLA..AGNGNNSRA.AYC TYNATDWEK.AL.KQTAE...RY..LELV...NN...TSNVT.TMF.NRSSGGDAEVTHLHFNCHGEFFYCNT
O_CM.ANT70        --M-A-D-LEG----------L--- ----QIDIQE..MR-..----..-----IG..GTA--S---.--- K------G-.I-.-----..--..----..--...-GSIN.MT-.-H-----L----------------
O_CM.MVP5180      --------TE-A--------TP---- -I-E-IAE--.D-YT..---R..-R--T-KRSNNTSPR--V.--- ---K-V--N.--.Q---I...--..-N--...-Q...-E---.II-.S-T-------S---------------
O_ES1.1158        V--I-R----------------L--- -S--EMVVQE..M--..----..----A-...GKEN--T--.---
O_FR.CFBCF01      ---------KT-E-------VS--I- -H----LS--.-M--..--LS..--------..-NSSIK--V.--- N-ST-E-T
O_FR.CFBCF02      --M-A-------Q-------T----- -Q---H----.--R-..----..--------..----SE--R.--- E--T-N-I
O_FR.CFBCF03      ----A---TNT-N----------I- -N---RGIK-..-G-..---S..V--GS--.DL-GN----I.--- D-DI-K-N
O_FR.CFBCF06      -----R--T-NT--------TS---- -M-K-RGKI-.R-AT..--LR..-V--AAK.TESQNTG--I.--- M--N-E-I
O_FR.CFBCF07      --L-A-------Q-----X-T--X-- -H----LK--.----..----..--------..----SE--K.--- X--X-X-V
O_FR.CFBCF08      --------RENA------------- ----EGNLTIQ--HS..----..---L--K..RNTTVR--S.-H- K--T-N--
O_FR.DUR          ---------SNSG--L---------- -V----NS--.----..----..----QIE.RE-K-A---T.-F- -------R-.T-.QGI--...--..----...-K...--PTE.I--.-K-N-----I-R----------
O_FR.VAU          -T----------E--LI---TN-TIA --------I-.K-MA..----..----A-S..NTK-D.T--.--- N-S----N-.--.-NIT-...--..----...EY...NQTDV.--KFGNH--E-----NFF----------
O_GA1.VI686       --------------------S-S-- ------H---.-M--..----..------E...E-KT---R.--- R--------.--.--MT-...--..----...--...---K-V--IF-Q-ND--P----------------
                                                  V3 LOOP
                  ^^^          ^^^   ^  *^   ^^^                  *   ^^^                          ^^   ^     ^^^           *       *
U_CONSENSUS_96    IIIRSENITNNAKTIIVQLNESVQIN CTRPNNNTRK.SIRI..GPGR..AFYTTGD..I.I.GDIRQ.AHC NISRTEWNN.TLQQVAKKL..RE..QFN....NK...T.IIF.NPS.SGGDP.EITTHSF.NCGG.EFFYCNT
U_AR.20021        ..---Q--SD-X-----HX---I--- ---------X.X-H-..----..-----X..X-.-.-----.--- -L--XX--X.--K-IVR-X..--..-XG....--...-
U_BR3.RJI0        -----Q--SD---I---H-----E-- ----------.--N-..----..---A---..D-.-.-----.-Y- ----AQ---.--E-I-I--..G-..--K....--...K.-A-.TQ-.-----T
U_BU1.91BU009     -------L-----I---------T-- ----YK-I-R.GTH-..-L--..-Y----..VE...-KKGP.P-- K-NG-G--K.I----T.-KL.-D..L--....--...-.---.--P.----L.-------.----.-------
U_CF1.4040        A------L-S-H--XXYSF---I--- ------IRNIRQRTHI.-S-Q..-IF--...KV.-.----K.-Y- T-NA-K--K.--H---M---..KS..LL.....--T..-.---.NQT
U_CF1.4056        ..--TK--SD-T-N-----KTP-N-T ---------T.--HL..----..---A---..D-.-.-----.--- -----D--K.--H--VRQ-..T
U_CF1.408         ---------D---I---H-----E-- ----------.-V--..---Q..T--A---..DM.K.-----.--- -V---K--Q.----I-TQ-..-K..Y-G....-TT..-.-N-.SKT
U_CF1.4087        VM----------------DSP-G--- ----G-T--R.-M--..----..T--.-.....-.----K.--- IVNG---K-.---K--NQ-..--..FT
U_KE.K124         VK---Q-L-----------VKP-R-- -------I--.----..---Q..--FA---..D-.-.-N---.-Q- -VNG----E.----ISTQ-..KK..H-M....--...-.---.---.-----.-------.--R-.-------
U_KE.KEN976       -M-------D---N----FT-PI--- ----------.---F..---Q..-----S...D-.-.---N..... .........------IA-..-RQ..--.....--...S.---.NS-.----I.-----T-.----.-------
U_NL.RW94028      ----------V--------QT-K-- ----S-----.-----..---Q..V--AN...E-.V.-----.--- -V--G--TQ.--RE--T--..--..Y-.....-N...-.TS-..T
U_SE.KI4803       VV----DF-----I-----K---E-- ----A--IKR.R-MHI.----..---K---..GM.T.-----.--- ----EK---.--K-IV.-KL.G-..--G....S-...-.---.NQ-.------.--VM---.----.-----KS
U_UG1.92UG035-2   -------L-----I-----KK--T-K -I--Y----E.GTH-..-----..-LF---...D-.-.-----.--- --------E.-------.-KL.GD..L-.....--E..E.---.--F.-----.-------.--K-.-------
U_UG1.C6080-09    V------L--------------E-- ----STKK-Q.TTP-..-L-Q..-L---R..N-.G.----K.--- --TKAA--K.-------.-KL.GD..LL.....-RT..-.---.---.-----.-------.----.-------
U_ZM.ZAM184       VM-------D---N----FT-P---S ----S----R.---F..---Q..-----N...D-.-.-----.-Y- S-NE-K--A.---K---.-QL.QN..Y-P....--...-.-S-.AN-.----L.-------.----.-------
U_ZR.Z3           V--------------------T-K-- ----GSDKKI.RQS-RI---K..V--AK-..G-.T.-...-.--- --TDG--R-.------IA-..-R..---....--...S.---.NS-.----I.-----T-.----.-------
```

Sequences with unusual inserts:

```
                                                                                      V3  LOOP
                          ^^^                 ^^   ^    ^ *^       ^^   ^                                              * ^^^     ^^^                              ^^    ^                     *
ALL_CONSENSUS  IIIRSENITNNAKTIIVQLNE..SVEIN CT.RP.NNN..T.RK.SIR...I..GPGQ..AFYA....TGD.I....IGDIRQAHC NISRTEWNNTLQQVAKKL.REH..F.....NK...T.IIFAPS.SGGDLEITT..HSFNCGG.EF
A_CF1.11423    VK-----F-----------DS..P-R-- --.--.--Y..-.--.M-P...T..----GQVI--....--K.-....-----K-Y- ---KR--.K-VKKGSY-T.-SK..-.....-
A_CF1.1189     V------------------DS..P-G-- -I.--.G--..-.--.---IMH-..---R..P---....R.G.V....--------- T-NKAK--Q---K--.-KL-GI..-......--...H.-.-TK
A_CF1.1286     V------------------DS..P-G-- --.--.G--..-.--.---IMH-..---R..S---....K.G.V....--------- T-NKAK--Q---K--E--.-GI..-N.....--...-.---TNH.------V--..--------.-V
A_CI2.CI326    VV-----------------SK..P-R-- --.--.G--..-.--.-W-SWH-..---R..---T....-DG.-....-----K-Y- -V-K-V-KD--K---RQ-.-KY..-N....-N...-.----NH.----V----..--------.--
A_CI2.CI47     -V-----------------SK..P-K-- --.--.G-Y..-.--.-V-T.G-..----..T---....--K.-....--------- -V--RD---------T--.--Y..-......--...-.---DK-.----V----..--------.--
A_UG.U455      -R-----F-----------VN..P-K-- -S.--.Y-...-.--.N--RYS-..-S--..---V....--K.-....--------- -V--RD--R-I----EQ-.KKK..-N.....--...-.----S-.----I----..--------.--
A_UG1.W2UG029  ......--SD---------TK..P-T-- --.--.---..IS-R.R-H...-..-R--..----.....--G.M....T-N----Y- -V-GSD--T-----VTQ-.-R-..WQ....-T...-.-V-KS-.-
B_AU2.C18CG    VV---D-F-----------SK..----T -V.--.---..-.--.--S...M..---R..---T....-E..-....-------Y- ---KAN-TD--E-I-R--.--Q..-E....--...-.-V-K--.----P--V-QFY------.--
B_BE.SIMI84    VV-----F-D----------K..--V-- --.--.--S..K.--.TLH...M..--KR..----....---.-GG.YT-Y------ ----AK--D--K-IV.-KLK-Q..-G...R--...-.-V-KQ-.----P--VM..--------.--
B_BR3.HSP204   -V-----L---I----I---K..--N-T --VV-.---..-.--.--P...M..-L-R..-L--....--E.-....--------- -L-SAD-----R-I-R--RE-K..-K....--...-.-A-KQ-.----P-
B_GB1.CPHL     VM---A-F-D---N-------..--V-- --.--.S--..-.-R.--S...-..---R..-W-T....--E.-....-----K--- -L--AD-----W-IV.-KL--Q..KQLQLE--...-.-V-NQ-S----P--VM..-T-----.--
B_GB3.CPHL7    VV---D-L------------A..----- --.--.---..-.--.G-H...-..---R..----....--E.-....--------- ----EK-----K-IVE--.--QEP-K....--...-.---NQ-.----P--VM..-----R-.--
B_GB5.4657     VV---A-FSD-T-S-------..--V-- --.--.---..-.--.T--TGH-..---R..SL-T....--Q.-....T----R--- ---G-K-----N-IVI--.--Q..-R....--...-.-V-NQ
B_HT1.D2HA599  V---A--FS-----------K..A-L-- --.--.---..-.-R.-VHSGH-..-G-R..TLFT....-.H.-....V----K--- -L-IAQ-----R---.-KL--Q..LG...D-R...-.---KQ-.----P--V-..---T---.--
B_HT2.H13968   V.--A--F------------K..A-A-- --.--.---..-.-R.-VHSGH-..-G-R..TLFT....-.H.-....V----K--- -L-IAQ-----R---.-KL--Q..LG...D-R...-.---KQ-.----P--V-..---T---.--
B_IN1.20016    VV-----F-D----------K..----- --.--.K-EI.K.-R.R-K...-..---R..--V-....--.T.....V--T---Q- -L--AK--E--K-IVI--.--Q..-
B_IN1.20023    VV---D-F-D----------K..----- --.--.--SI.K.-R.R-H...-..---R..--F-....-.N.T....V--T---Q- -L--AK--E--K-IVI--.-
B_IN1.30005    VV---D-F-D----------K..----- --.--.D-E..I.-RSRLQ...V..---R..--V-....A-.T.....A--T---Q- -L--AK-IE--KLIVI--.--
B_IT.PD        ...--A-F-----N-------..--K-- --.--.Y--..IKIR.H-H...-..---R..--H-.....-E.V....A-N------ ---GEK--D--N-IVI--.--Q..YG....-N...-.-
B_NL.X1con     VVV----F-------------..-I--- --.--.--Y..-.--.G--...-..---R..-VI-....-EK.-....--------- -L-----EK--K-I-N--.--QSQ-K....--...-.---NQ
B_NL5.pt12con  VV-----F-DD----------..--V-- --.--.---..-.--.G--IGH-..---R..---T....--K.-....-----K--- ----KD-----K-IVI--.--Q..-G....--...-.-T-NH
B_SK1.BTS15    VVL----F-------------..T-Q-- --.--.---..-.--.--H...M..---R..----....-DA.-....--N------ -L--E--GK--K-IVI--.EKQ..-NFTFE--...-.-T-KN-
B_TW1.84       X..------D---------K-..--N-- S-.--.-TE..I.-RRI-H...-..---R..--S.....-G.-....--------- -L-K-------AHIVPPX.I-Q..-G....--...-.-V
B_TW1.90       X..K-----D---------K-..P-N-T --.--.D-T..I.K-RI-H...-..---R..---T....R.K.-.....-.LM----. KL-K-K-----A-IV.-KF--Q..-G....--...-.-V-
B_US.PD        ...--A-F-----N-------..--K-- --.--.Y--..IKIR.H-H...-..---R..--H-.....-E.V....A-N------ ---GEK--D--N-IVI--.--Q..YG....-N...-.-
B_US.RJS       V.-----F-------------..-IV-- --.--.Y-T...K-..--HIH-..---R..----....--G.-....M-------- -L--A--ES--K---T--.--Q..-G....--...-.-V-KQ-.----P--V-..--------.--
B_US10.SFBUcon VV---D-FSD-----------..--V-- --.--.---..-.--.G-Y...-..---R..-V--....-ER.-....-----K-F- ----AK-----K-IVR--.--QDQ-K....--...-.-V-KQ-.----P--VM..-M---R-.--
B_ZA.ZA510     VVV----F-------------..-I--- --.--.---..-.--.---IGH-..---R..-L-T....--R.-....--------- -----K-----N---T--.K-Q..-G....--...-.-V-Q--.----P--EM..--------.--
C_IN3.IND5     --VS---L---------H---..----R -A.--V---..-.--.---..-..----..--F-....---.-....--------- ----G---K---R-SE--.A-L..-P....--...-.-K-NA-.----------..-----R-.--
D_NL2.ZR94022  -------L-D-V---------..--A-- --.--.Y--..-.-Q.-TN...-..----..----...T-KR.-....--------- -V------K--RW--R--.GRL..L......--..T-.-T-
D_TZ2.023      -------L-----N-----K-..--S-- --.--.YR-..-.-R.RTP...-..-L--..-L-T..TNIN.-....L----R-Y- ----AR--R------G--.GNL..L......--..T-.---K-P.----P----..--------.S.
D_TZ2.053      -------LA----I-------..--T-- -S.--.YIKN.I.-R.RTH...-..-S--..-Y-T....-.N.-....G-T--P--- ----A---K-I----E--.GDL..L......--..T-.---K--.----P----..-------X
D_TZ2.064      -------L-----N-------..--T-- --.--.YRSS.-.-R..TP...-..-I--..SL-T....-KN.-....M-------- ---G-S--K-------.-KLGDL..L.....-Q..T-.---K-P.----P----..-X-----.S.
D_TZ2.080      -----K-L-----I-------..--T-- --.--.YHKD.-.IR.RTH...-..-S--..-L-T....-.R.L....K------Y- ---GIG--K-------.-KL-DL..-.....-Q..T-.---K-P.----P----..--------.S.
D_TZ3.1627     ......-L-----I-------..--T-- --.--.YK-Y.-.I-.KTH...-..-S--..-I-T....A.G.-....Q-I-ER-Y- ---GA--SK------T--.-DL..-......--..T-.-N-K--.----P----..--------.--
D_TZ3.4622     -.----SL-----I-------..--T-- --.--.YIK..Y.-Q.KTS...-..-Q--..-L-T....-.R.LEPKPT-K-EP-Y- ---GA---K-I---T.-RLKHL..-P....--...-.-N-Q--.----P----..--------.--
D_UG.U44342    -------L---V-------K-..--A-- --.--.---..-.-Q.-TQ...-..----..-LFTKRVT-.R.-....--------- ---GAK--K-------.-KLGGL..L.....-Q..T-.---K--.----P
D_UG1.W2UG00   ......-L------L------..----- --.--.YY-Q.I.-Q.RTS...-..-Q--..-L-T....-.R.V....T----K-Y- ---KAG--K-------.-KLGDL..-.....-Q..T-.---K--.-
D_UG1.W2UG038  ......--S------------..--T-S -A.--.FYA..IE-Q.RTP...-..-Q--..VL-T....-.K.K....--R-G---- --TGAD-KK-------.-KL-DL..LN....IT...-.VN-K--.-
D_UG1.W2UG040  ......-L-----I-------..--T-- --.--.---..-.-Q.-TH...-..----..-I-TLY.T-.K.V....V-------- ----AA--K-------.-KLGDL..L.....-Q..T-.-V-K
D_UG1.W2UG053  ......-L-------------..-IP-- --.--.YEKI.K.-R.GTP...-..-L--..-YLT....-.Q.-....T-Y------ ----G---K-------.-KLGDL..L.....-Q..T-.-V-K
D_UG1.W2UG065  ......-------N-------..--T-S -I.--.Y-K..-KIQ.RTS...-..-L--..-L-T....NDK.-....--N----Y- -V--AG--K-------.-KLGDL..L.....-Q..T-.---Q
D_UG1.W2UG070  ......-L-----I------K..----R -A.--.YIK..I.WR.RTH...-..-S--..-YSTKR.........QNYTGP--- ---GGD--D--K---E--.-HL..L......--...-I---K
D_UG2.1665     -------L-----I-------..--T-- --.--.YK-I.-.IQ.RTP...-..-L--..-L-T....-.K.R....--V-G--S- ---GAR--K----
D_UG2.1685     -------L-----I-------..--A-- -S.--.YR-V.-.IQ.RTS...-..-S--..-L-T....-.K.R....--Y-K---- ----A---K-------.-
D_UG2.653      -------L-----V-------..--T-- --.--.YKS..I.-R.........-..---RWQTY-T....-.N.-....T-...R--- ----AG-KK-------.-KL--L..L......--..T-.---K--.----P
D_UG2.G        -------L-D---V-------..--T-- -I.--.Y-KR.-.-Q.RTH...-..-S--..-L-T....-.K.-....V-------- --------K-------.-K.
D_UG2.G2       -------L-----------K-..--A-- --.--.YD-..IK-Q.RTP...-..-Q--..-L-T....-.R.L...TTRR-G-P-- ---GA---K-------.-KLGDL..L......--..T-.-S-K--.----P
D_UG3.70       -V-----L-----I-------..--I-- --.--.Y--..IKIQ.RTP...-..-R--..-LFT....-.R.-....K-IKG---- ---KAG--K-------.-KLQD
D_UG5.UG266    -V-----L-----I-------..--I-- --.--.Y--..IKIQ.RTP...-..-R--..-LFT....-.R.R....K-IKG---- ----EG--K-------.-KL-DL..I......--..T-.---K--.----P----..--------.--
D_UG6.980      ......-L-----I--------..P-I-- --.--.YQK..VHTY.R.P..H-..-.-R..-S-T....-.R.-....T-N------ ----A---K------N--.-G
D_UG7.963      -------L-----I-----K-S.--P-- --.--.Y--..-.-Q.GTH...-..---R..---T....-.E.-....--------- ----VG-DK--H---G--.-AF..LK....TP...P.---R--.----P----..--------.--
D_ZR.JY        -------L---V-----H---..----- --.--.D-KI.-.-Q.-TP...-..-L--..-L-T....-.R.-....K------Y- ---AAA--K-------.-KLGDL..L.....-Q..T-.---K-
F_RO2.RM53012  .....Q--SD-T-----H---..T-Q-- --.--.---..-.--.G-H..I-..-V-R..P--R....-V-.-....V----K--- -V-GAQ----V-R-
G_NL.127C      .M-----F-D-T-V-------SET---T -I.--.---..-.--.--T...F..----..----....-SN.-....----K--Y- -VNK-K--QM-LK-KTH-.Q-K..YP....-G...-.-Y-NS
G_NL1.UM92101  -M-----F-D-T-V------KNET---- --.--.---..-.--.--T...F..----..----....-SN.-....--N----Y- -VNK-K--RM-RK-QTQ-.K-K..Y.....--...-.-Y-
H_CM.CA13      VM-------D-T-N-----KN..P-V-- --.--.---..-.-R.-M-.IG-..-R--..T-H-....I-A.-....--------- ---GAK--DM-TK--TQ-.GK-..L......--..SK.-Q-K--.----I--I-..-M---A-.--
```

Sequences with unusual inserts (continued):

```
                          *
ALL_CONSENSUS  F...YCNT
A_CF1.11423
A_CF1.1189
A_CF1.1286     -LLQ-I
A_CI2.CI326    -...---
A_CI2.CI47     -...--Y
A_UG.U455      -...---
A_UG1.W2UG029
B_AU2.C18CG    -...---S
B_BE.SIMI84    -...---S
B_BR3.HSP204
B_GB1.CPHL     -...----
B_GB3.CPHL7    -...----
B_GB5.4657
B_HT1.D2HA599  -...----
B_HT2.H13968   -
B_IN1.20016
B_IN1.20023
B_IN1.30005
B_IT.PD
B_NL.X1con
B_NL5.pt12con
B_SK1.BTS15
B_TW1.84
B_TW1.90
B_US.PD
B_US.RJS       -...---.
B_US10.SFBUcon -...----
B_ZA.ZA510     -...---S
C_IN3.IND5     -...----
D_NL2.ZR94022
D_TZ2.023
D_TZ2.053
D_TZ2.064
D_TZ2.080
D_TZ3.1627     -...----
D_TZ3.4622     -...----
D_UG.U44342
D_UG1.W2UG00
D_UG1.W2UG038
D_UG1.W2UG040
D_UG1.W2UG053
D_UG1.W2UG065
D_UG1.W2UG070
D_UG2.1665
D_UG2.1685
D_UG2.653
D_UG2.G
D_UG2.G2
D_UG3.70
D_UG5.UG266    -...---.
D_UG6.980
D_UG7.963      -...S--.
D_ZR.JY
F_RO2.RM53012
G_NL.127C
G_NL1.UM92101
H_CM.CA13      -...----
```

A Subtype

At this time there are viral sequences from 207 HIV-1 infected individuals associated with HIV-1 subtype A. The A subtype consensus sequence (A_CONSENSUS_96) generated from these sequences was based on the most common amino acid found in each position in an alignment of these sequences. All of these sequences have been published and/or have been made available for printing in the database by their authors.

1) **BJ1.ID#:** These 18 sequences are from female prostitutes, born in either Ghana or Togo, who live in Benin. 15 are from directly sequenced PCR products, derived via RT-PCR from patient serum RNA. 3 (233, 251 and 253) are from cloned PCR products, also by RT-PCR from serum RNA. Another subtype A sequence (366, U61870) is not included here, because it was nearly identical (254 of 255 bases identical) to the SF170 (M66533) sequence from Rwanda isolated in 1988, and thus it likely represents a lab artifact. [Heyndrickx et al.(1996)]. GenBank accession numbers U61854–U61869, U61871 and U61873. Two subtype G isolates were also found in this study.
2) **BY.BLR10A:** This sequence is from Byelorussia. [Lukashov et al.(1995)]. GenBank accession number L38411.
3) **CA.HWCL1:** This sequence is the first published sequence of subtype A HIV-1 in Canada. The patient had moved from Uganda in 1983, and was diagnosed as HIV+ in 1989. Viral RNA was recovered from archived, stored patient serum by binding viral particles to CD4-coated wells of an ELISA plate. The RT-PCR amplification product was cloned and 10 clones were sequenced. It is not clear whether this sequence is from a single clone, or a consensus of all 10. [Montpetit(1995)]. GenBank accession number U34049.
4) **CF1.ID#:** These thirteen sequences are from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990–1991. The sequences are consensus sequences from cloned PCR products. DNA was isolated from co-cultured PBMCs. [Murphy et al.(1993)] GenBank accession numbers L11457–L11458, L11461–L11463, L11469–L11471, L11474–L11475, L11477–L11479, L11484–L11496, L11498, L11518 and L11523–L11524. Some of them were sequenced again by Schmitt et al. unpublished: U43275, CF1.4023; U43109, CF1.4054; U43139, CF1.11423; U43136, CF1.1286.
5) **CF2.GAN and SAS:** These sequences were kindly provided prior to publication by Dr. MP Kieny of Transgene, Strasbourg, France. They are part of a set of 14 HIV-1 gp120 isolates from Bangui, Central African Republic. The gp120 proteins have been expressed in a hybrid gp160 vaccinia virus expression system. Year of isolation and health status of individuals from which the virus was isolated were not provided. Viruses were isolated in the CAR by C Mathiot and B You (Pasteur Inst., Bangui), grown by F Barre-Sinoussi and A Deslandres (Pasteur Inst., Paris), and cloned and sequenced by D Schmitt and MP Kieny. GenBank accession numbers U43111, U43171, M81063, M81064.
6) **CI1.ID#:** These six sequences are from six different AIDS patients suffering from pulmonary tuberculosis at the Pneumology Hospital of Cocody, Abidja, Ivory Coast. In this study sequences from viral RNA from viruses from serum cocultured on donor PBMCs, proviral DNA sequences from each patients' PBMCs after coculture with donor PBMCs, and proviral DNA sequences directly from uncultured PBMCs were determined. For the cocultured samples viral RNA was harvested from the culture supernatant. PCR or RT-PCR was used to amplify the env V3 region, and 4-7 cloned PCR products were sequenced. A total of 66 sequences from the six patients were published. All 66 were subtype A, and intrapatient sequences were more similar than interpatient sequences. [Audoly et al.(1996)]. GenBank accession numbers U59559–U59624.
7) **CI2.CI-ID#:** These sequences are from 11 of 13 isolates from individuals from Abidjan, Cote d'Ivoire. CI-14 and CI-20 were symptomatic, and the others were asymptomatic. CI-14, CI-45 and CI-47 were serologically dually reactive for HIV-1 and HIV-2. The C2V3 region is part of a 900 bp sequence. Samples were collected between May 1990 and Sept. 1991. Virus was cultured with donor PBMCs, nested PCR amplified, 3–4 clones were sequenced, and the consensus of those clones is presented here. [Janssens et al.(1994a)]. GenBank accession numbers X72024–X72027, X72030–X72039, X72043–X72056, and X72059–X72065.
8) **CM1.CA-ID#:** These sequences are 11 of 17 sequences from a very diverse set of isolates from Yaounde and Douala, Cameroon. The sequences were derived from asymptomatic (CA7, CA11, CA15, CA17, CA18, and CA21) and symptomatic (CA1, CA2, CA6, CA19 and CA22) individuals. Virus was isolated

by culture with donor PBMCs, and nested PCR amplified. A single clone was sequenced representing each HIV-1 isolate. [Nkengasong et al.(1994)]. GenBank accession numbers for the entire set of 17 envelope sequences: X80438–X80454.

9) **CY.HOcon:** This is a consensus sequence from four individuals who were epidemiologically linked to one another: a father, a mother, their child and a woman who was a heterosexual partner of the father. These samples, like others in this study (see also subtypes B, C, F and I) were collected in February 1994 from the AIDS clinic in Nicosia, Cyprus. PBMC DNA was PCR amplified and cloned. Individual clones were sequenced. The father (patient HO34) was 36 years old and asymptomatic. He was known to have been seropositive for at least 2 years and had a CD4 count of 410. The mother (patient HO17) was 35, had been infected for at least two years, had a CD4 count of 2 and died in May 1994. The child (patient HO49) was 2 years old, had been infected since birth, and was asymptomatic, with a CD4 count of 1,211. The partner of the father (patient HO42) was 45 years old, had been infected for at least 4 years, was symptomatic with CD4 count of 80, and died in August 1994. Because of the close epidemiological linkage, a consensus of all 5 sequences is presented here. [Kostrikis et al.(1995)]. GenBank accession numbers U28683 (child); U28674, U28719 (father); U28677 (father's partner); U28665 (mother).
10) **DJ.DJ-ID#:** These three sequences from Djibouti were from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession numbers L22939, L22941, and L23064.
11) **GA.VI191:** This sequence from Gabon was from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession number L22952.
12) **GA1.ID#:** These 6 sequences are from Gabon. Two (G41, G135) are from 1988-1989 samples from patients with AIDS living in Franceville Gabon. VI685 is from a 1992 sample of an AIDS patient from the Libreville General Hospital. VI1076 is from a 1993 sample of an AIDS patient from the Libreville General Hospital. LBV2310 and LBV23 are from 1988 samples from asymptomatic individuals sampled from the general population. Method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced. [Delaporte et al.(1996)]. GenBank Accession numbers X90914, LBV23; X90915, G135; X90917, G41; X90918, LBV2310; X90921, VI1076; X90924, VI685. See also subtypes C, D, F, G and O sequences from this same study.
13) **GH.D687:** A single sequence from an individual from Ghana. Virus was cocultured on PBMCs and the env gene was PCR amplified. Provided by Georg-Speyer Haus, Frankfurt, Germany (Dr. Ursula Dietrich). [Dietrich et al.(1993)]. GenBank accession numbers L07652, X68407.
14) **IN1.ID#:** These two entries are from 1992 dried blood spot samples from Vellore near Madras, in Tamil Nadu state in southern India. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. Samples came from previously identified HIV seropositive homosexual men. Other samples from the same region were all subtype C (see C_IN4.#). [Cassol et al.(1996)]. GenBank accession numbers U53286 and U53291.
15) **KE.K89:** This sequence is named "KENYA" in the GenBank entry, but is identified as K89 in the original manuscript. It is a Kenyan sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession number L22943.
16) **KE.KEN-ID#:** These 19 sequences were derived from patients who were part of a 1990–1992 cohort study of maternal risk factors in mother to child transmission, including 22 pregnant women and an infant from Kenya. The C2V3 region was sequenced. [Janssens et al.(1994c)]. GenBank accession numbers for the entire set of 23 patients surveyed in this study: U12984–U13006.
17) **KE1.-ID#:** These 6 sequences were derived from patients who were part of a May-June 1992 study of pregnant women from the Pumwani Maternity Hospital in Nairobi, Kenya. Viral RNA was concentrated from patient serum just prior to delivery, and the envelope C2-V3 region was amplified by RT-PCR. The PCR product was cloned and 20 clones from each patient were sequenced. Two other patients from this

study had viral subtypes C and D. [Zachar et al.(1996a)]. Genbank accession numbers U32658, U33763, U33764, U33766, U33767 and U34905.

18) **NG.IBNG:** An envelope gp120 sequence from Nigeria was kindly provided by Dr. Tom Howard from the University of Southern California (USC) [Howard et al.(1994)] and the complete genome has now been sequenced [Howard & Rasheed(1996)]. Genbank accession numbers U48628 and L39106.
19) **NL1.ID#:** These 13 sequences are from recent immigrants to The Netherlands from various countries. The fist two letters of the ID# represent the two letter country code for the previous residence of the patient. The next two numerals represent the year of isolation. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced. [Lukashov et al.(1996)]. GenBank accession numbers L76877, GH902304; L76903, GH9401488; L76883, GH915200; L76913, GH9501283; L76893, GH929927; L76896, ZR9303337; L76870, RW890388; L76875, RW893568; L76905, TZ9401664; L76881, KE913514; L76891, UG928308; L76887, AO924646; L76889, TZ925825.
20) **RU1.ID#:** These 4 sequences are from Russia. Bobkov et al. 1996 Unpublished. GenBank accession numbers: U33098, IVA6; U33097, GAN1; U33104 U33105, MLY10; U33106 U33107 U33108, SHL9.
21) **RW.564C** This sequence represents 10 identical sequences generated from PCR amplified plasma RNA from one of three infants in a Dutch mother/infant study. Patient pair 564 was from Rwanda. A sample was collected from the infant at 30 months of age. Samples were also collected from the mother at 12 and 30 months after the birth. Mother sequences are not included in this consensus. [Mulder-Kampinga et al.(1993)]. [Mulder-Kampinga et al.(1995)]. The child 564 Env sequence is from the entry with GenBank accession number Z47881. Mother 564 sequences are in entries with GenBank accession numbers Z47882–Z47902. Mother sequences are not included in this alignment. Gag gene sequences from mother/child pairs are also available in Genbank accession numbers Z47903–Z47911, Z47912–Z47928, Z47929–Z47935 and Z47936–Z47950. The second mother/child pair was from the Netherlands, see G_NL.127C. The third mother/infant pair in this study was from the Netherlands, see B_NL.114C.
22) **RW.SF1703:** This sequence is from Rwandan isolate sf170, a biologically active clone reported to be macrophage-tropic. [Cheng-Mayer et al.(1988)]. GenBank accession number M66533. See also U61870, which is not reported to be from SF170, but is greater than 99% identical to it.
23) **RW1.W2RW-ID#:** Seven sequences from asymptomatic individuals from Rwanda sampled in 1992. Consensus, PCR-clones, cell-culture DNA and RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. Relevant papers are: [De Wolf et al.(1994)]; [Osmanov et al.(1994)]; [Gao et al.(1994a)]. GenBank accession numbers U08630–U08640, U08645, U08647–U08665, U08763–U08766, and U08793–U08794. One of these sequences may be mislabelled in the database; U08634 is labelled as patient 016, but is identical to sequences from patient 008. Sequences from samples 92WR020 and 92RW021 were greater than 98each other, so only 92RW021 is presented here.
24) **RW2-ID#:** Nine consensus sequences from Rwanda. Saah, A. Unpublished 1994. GenBank accession numbers U23216–U23373.
25) **RW3.PV#:** Bex,F. et al. Unpublished. These five sequences are from Rwandese patients with AIDS. The sequencing was done by the EEC Centralized Facility for HIV genome characterization, Georg-Speyer-Haus, Frankfurt, Germany. The complete envelope gene for PVP1 is available from a clone obtained after short-term co-culture on donor PBMCs. Two other shorter sequences of PVP1 env direct from patient PBMCs are also available. GenBank accession numbers L07082–L07091. L07088 and L07089 were withdrawn from GenBank by the authors, who felt they may represent PCR artifacts.
26) **RW4.#:** Kampinga,G.A. et al. Unpublished. These 8 consensus sequences and 2 individual sequences, are from 7 infants and 3 mothers in a study of mother-infant transmission in Rwanda. Mother 566 was apparently infected with both subtype A and subtype C HIV-1, but the env regions from the child clones were all subtype A. See also C_RW1.ID#. GenBank accession numbers for child 566 are Z76160–Z76161, Z76167–Z76168, Z76169–Z76176, Z76233–Z76248, Z76262–Z76273 and Z76717–Z76724; mother 730, Z76353–Z76362; child 564, Z76074–Z76083; mother 226, Z76046; child 538, Z76134–Z76143, Z76373–Z76382, Z76393–Z76412; child 074, Z75958; child 082, Z75998–Z76009, Z76650; child 081, Z75959–Z75968; child 618, Z76198–Z76207; mother 439, Z76048–Z76057, Z76064–Z76068 and Z76070.

27) **TZ.TAN-ID#:** These four sequences are from a set of 14 Tanzanian samples from symptomatic individuals, using serum samples taken in 1988 to generate PCR clones from viral RNA for sequencing. [Zwart et al.(1993)]. GenBank accession numbers L01313, L01315–L01316, L01335, L01337–L01339.
28) **TZ2.ID#:** These two sequences were from patients at a clinic in Dar es Salaam, Tanzania. The individuals from which the virus was cultured showed clinical signs of AIDS, and the year of viral isolation was 1988. Viral cDNA was PCR amplified from donor PBMC, and one cloned PCR product per donor was sequenced. [Siwka et al.(1994)]. GenBank accession numbers U12408, U12409.
29) **TZ3.ID#:** These 4 sequences are from the Mara region of rural northwest Tanzania [Robbins et al.(1996)]. Subtype D was also found in this study. GenBank accession numbers U61875–U61878.
30) **UG.1033:** This sequence is a consensus sequence of blood and CSF samples taken from a Ugandan patient 1033, CDC class IV-A. [Keys et al.(1993)]. GenBank accession numbers Z23177, Z23182–Z23184, and Z23220–Z23223.
31) **UG.92UG037:** This sequence is from a complete genome PCR amplified from proviral DNA. The patient was a 31 year old asymptomatic female from Entebe, Uganda. [Gao et al.(1996b)]. GenBank accession number U51190.
32) **UG.964:** A single sequence used in a study of the impact of sequence variation on the distribution and seroreactivity of linear antigenic epitopes. The sequence was derived from PCR amplified DNA from peripheral blood leukocytes. The patient was an asymptomatic individual from Uganda. [Pestano et al.(1995)]. GenBank accession number U11599. See also B_US17.ID#, C_UG1.45, and D_UG7.ID#.
33) **UG.U455:** This sequence is from the 1985 Ugandan isolate U455; the complete genomic sequence is available. [Oram et al.(1990)]. The env ORF in this sequence is interrupted by an in-frame stop codon beyond the COOH end of the V5 region. GenBank accession number M62320.
34) **UG.UG06:** This sequence is from blood collected from the Mulago Teaching Hospital in Kampala, Uganda. Viral RNA was harvested after 10-14 days of coculture with donor PBMCs and reverse-transcribed with AMV-RT. The env V3 region was pCR amplifed and cloned. This sequence is from an individual clone. [Atkin et al.(1993)], [Pestano et al.(1993)]. GenBank accession number M98503.
35) **UG1.W2UG-ID#:** Three sequences from asymptomatic individuals from Uganda in 1992. Consensus, PCR-clones, cell-culture DNA and RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. Relevant papers are: [De Wolf et al.(1994)]; [Osmanov et al.(1994)]; [Gao et al.(1994a)]. GenBank accession numbers U08666–U08669, U08767–U08770, U08788–U08792, U08795, U09124, U09125 and U09127
36) **UG2-ID#:** These 11 sequences are part of a set of sequences derived from 22 Ugandans who were attending an AIDS clinic, sampled in 1990. Consensus, PCR-clones, peripheral blood DNA. [Albert et al.(1992)]. GenBank accession numbers M98902–M98905, M98908–M98910, M98914–M98917, M98919, M98924–M98928, M98938–M98941, M98946–M98966, and M98976–M98978.
37) **UG4.UG-ID#:** Two Ugandan sequences from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession numbers L22957 and L22951.
38) **UG5.ID#:** These 8 sequences are from Gulu, northern Uganda. They are from direct sequence of PCR product amplified from uncultured PBMCs. Blood samples were drawn from 217 pregnant women attending a clinic in Gulu, northern Uganda. Ages ranged from 17 to 37 years. The 29 seropositive women (13.4% of the 217 tested) were all asymptomatic. [Buonaguro et al.(1995)]. Genbank accession numbers U44878–U44880, U44882, U44883 and U44885–U44887. Two subtype D sequences were also found in this study (see D_UG8).
39) **ZR.Z321:** This sequence is from the 1976 Zairean isolate Z321. [Srinivasan et al.(1989)]. GenBank accession number M15896.
40) **ZR1.ID#:** These ten sequences are part of a set of 14 A and D sequences from women from Zaire; 8 were healthy, 4 showed minor signs of illness, and 2 had AIDS. PCR-direct, peripheral blood DNA. [Potts et al.(1993a)]. GenBank accession numbers L19624–L19626, L19628–L19630, L19632–L19634, and L19636.
41) **ZR3.ID#:** These sequences are from Zaire. They are unpublished, by M. Reitz et al. GenBank accession numbers U43097–U43100.

B Subtype

At this time we have included viral sequences from 975 HIV-1 infected individuals associated with HIV-1 subtype B. The B subtype consensus sequence (B_CONSENSUS_96) generated from these sequences was based on the most common amino acid found in each position in an alignment of these sequences. Please note that none of the studies which have published sequences of only the V3 loop sequences are included here, as the DNA sequences were deemed too short for phylogenetic analyses. (For example, LaRosa G, et al., *Science* **249**:932–935 (1990) and Fouchier RAM, et al., *J. Virol.* **66**:3183–3187 (1992).)

1) **AR.21281:** This sequence is from direct sequencing of PCR product from uncultured PBMCs, from a 1993 sample from Buenos Aires, Argentina. The patient had AIDS and reported promiscuous heterosexual risk behavior. Two other samples taken from unrelated patients in 1993 were subtypes F and one was found to be a subtype B/F recombinant. [Marquina et al.(1996)]. GenBank accession number U68525.
2) **AR1.ID#:** These 11 sequences are from Rosario, Argentina. A total of 24 patients from different risk groups visiting a clinic in Rosario were included in this study. Of the 14 sequences determined, 11 were found to belong to subtype B and 3 were found to belong to subtype F. DNA was extracted from whole blood and PCR amplified. PCR products were directly sequenced. Subtypes of all 24 patients were tested by HMA. [Campodonico et al.(1996)]. GenBank accession numbers U37030, U37031 and U37034–U37042.
3) **AU1.ID#:** Put forth as evidence that coinfection by multiple HIV-1 strains can occur in vivo, these three concensus sequences (MRC1,MRC2, and MRC3) come from an Austrailian homosexual male who had been infected by more than one sexual partner, and harbored three distinct strains of HIV-1 B. The authors also found recombinant sequences, not included here. The sequences were PCR amplified from plasma RNA and PBMC DNA. [Zhu et al.(1995)]. GenBank accession numbers U16372–U16388.
4) **AU2.C18CG:** This sequence is one of a group relating to an Australian blood donor infected with HIV-1 and six Australian recipients, all of whom remain symptom free with normal CD4 counts 10 to 14 years after infection. [Deacon et al.(1995)]. Samples from only the donor, D36 (U37271), and two patients, C18 (U37267, U37270) and C98 (U37268, U37269), appear to have been sequenced. These sequences have deletions in the nef gene and in the region of overlap of nef and the U3 region of LTR. The authors point to the importance of NEF or the U3 region of LTR in determining the pathogenicity of HIV-1 and suggest this strain of HIV-1 as a possible basis for a live attenuated vaccine. The complete genome of the virus from recipient C18 is in the entry with accession number U37270.
5) **BE.SIMI84** One of two cloned env sequences from a patient with AIDS from Belgium. A vaccinia construct that expresses this gene was created to vaccinate the patient's non-infected brother with the goal of immune therapy by adoptive transfer of lymphocytes. [Bex et al.(1994)]. GenBank accession number L07421.
6) **BR.002con** This consensus sequence is from Genbank accession numbers L35489–L35493. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. Ranjbar,S. et al. Unpublished (1994).
7) **BR.BZ** This sequence is from an individual in a Brazilian HIV cohort study. PBMC DNA was PCR amplified in two sequential rounds, and six cloned PCR products were sequenced on both strands. A single clone containing an uninterrupted envelope open reading frame was reported. [da Costa et al.(1995)]. Genbank accession number U28336.
8) **BR1.ID#:** These 21 sequences represent the B env subtype sequences found among 22 Brazilian outpatients with varying degrees of disease progression. They are consensus sequences from cloned PCR products. PCR was performed on PBMC DNA. [Potts et al.(1993b)]. GenBank accession numbers for 21 of the 56 clones from which consensus sequences were calculated: L19225–L19237, L19240–L19246 and L20963.
9) **BR2.W2BR-ID#:** These 13 sequences are from individuals from Brazil. They are consensus sequences from cloned PCR products. Some of the clones were from cell-culture DNA, and some from cell-culture supernatant RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. [De Wolf et al.(1994)], [Osmanov et al.(1994)], [Gao et al.(1994a)]. GenBank accession numbers U08670–U08714, U08771–U08778, U08780–U08782, U08792, U08796, U08797–U08800.

10) **BR3.RJ- or SP-ID#:** These 32 sequences came from the Brazilian cities Rio de Janeiro and Sao Paulo. Some of the sequences that are very short, containing V3 loop fragments insufficient for phylogenetic analysis, are not included here (4 of the 36). The full set included 34 viral sequences of the B subtype, an F subtype and a B-F recombinant (see subtypes F and U). The year of isolation for the sequences ranged from 1990–1992 for Rio de Janeiro, and 1992 for Sao Paulo. The only two with CD4+ cells < 200 were RJ636 and RJ27. The CDC clinical class ranged from II-IV. DNA extracted from PBMCs of HIV infected individuals was amplified, and the PCR product was directly sequenced. [Morgado et al.(1994)] and [Sabino et al.(1994c)]. More recent unpublished sequences from the same isolates are also included here. GenBank accession numbers U00400–U00401, U00403, U00405, U00407–U00414, U00416–U00418, U00421, U00424–U00425, U00427, U08975, U31586, U31587 and U31589–U31591.
11) **BR4.BZ-ID#:** These 2 sequences are from seropositive Brazilian patients. Virus was cultured on donor PBMCs and proviral DNA was harvested from positive cultures. PCR was used to generate sequencing templates. [Louwagie et al.(1994)]. GenBank accession numbers L22087 and L22088. The gag gene sequences from these same isolates are also available in L11752 and L11754. See also F_BR2.BZ-ID#.
12) **BR5.ID#:** These 10 sequences are from entries with GenBank accession numbers L19328–L19337. Bandea, C. I. Unpublished (1993).
13) **BR6.ID#:** These six sequences are from patients living in cities in Brazil (P3, Sao Paulo; P4, P6 and P7, Bahia; P8, Parana; P9, Rio de Janeiro) and sampled between 1987 and 1989. Sequences were determined from directly sequenced PCR products (except P6 which was a cloned PCR product), after coculture of patient PBMCs with donor PBMCs. [Couto-Fernandez et al.(1994)]. Genbank accession numbers X78512–X78517.
14) **CH1.ID#:** These 19 sequences came from 24 individuals living in Geneva, Switzerland who were recently infected at the time of blood drawing. Samples were collected between January 1988 and September 1993. Sequences were determined directly from PCR products of uncultured PBMC DNA or serum cDNA. All subjects were asymptomatic, 19 subjects had p24 antigen levels ranging from 5 to 6,357 pg/ml and 5 subjects had no detectable p24 antigen. Two subjects were epidemiologically linked (K11 and K16) so only one of those two is presented here. Two other individuals showed identical DNA sequences over the entire V3 region (K53 and K77) so only one of them is presented here. Three other individuals (K13, K42 and P4) had sequences nearly identical to the LAI (IIIB) laboratory strain of HIV. Although the authors are convinced that these are not contaminants, and that a IIIB-like strain of HIV is circulating in Geneva, they are not included in this alignment. [Antonioli et al.(1995)]. GenBank accession numbers U10957–U10980.
15) **CI.CI-22:** A single B subtype sequence from a set of 13 isolates from individuals from Abidjan, Cote d'Ivoire. CI-22 was symptomatic. The C2V3 region is part of a 900 bp fragment that was sequenced for each individual. Samples were collected between May 1990 and Sept. 1991. Virus was cultured with donor PBMCs, nested PCR amplified, 3 clones were sequenced, and the consensus of those clones is presented here. [Janssens et al.(1994a)]. GenBank accession numbers X72040–X72042.
16) **CM.CA5:** A single B subtype sequence from a set of 17 sequences from a very diverse set of isolates from Yaounde and Douala, Cameroon. The sequences were derived from asymptomatic and symptomatic HIV infected individuals, specifically, patient CA-5 was asymptomatic. Virus was isolated by culture with donor PBMCs, and nested PCR amplified. A single clone was sequenced representing each HIV-1 isolate. [Nkengasong et al.(1994)]. GenBank accession numbers for the full set of 17 sequences which includes CA5 are X80438–X80454.
17) **CN1.-ID#:** These 54 sequences are form the Yunnan Province, China. No information is yet available. Unpublished 1995, Y. Shao and H. Wolf. Several of these sequences (U20009, U20012, U20013, U20018, U20023, U20024) are greater than 96GenBank Accession numbers U20001–U20054.
18) **CN.1798:** This sequence is from a dried blood spot collected in 1992 from the spouse of an IV drug user in China. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. [Cassol et al.(1996)]. GenBank accession number U53316.
19) **CU.94CU053:** This sequence is from a bisexual male, most probably infected via heterosexual contact in 1992, in Cuba. Virus was isolate in 1994, 2 years after seroconversion, by cocultivation of patient PBMCs with donor PBMCs. This isolate exhibits a rapid/high, syncytium-inducing phenotype. [Gomez et al.(1996)]. GenBank accession number U48855.
20) **CY1.HO#:** These sequences are from samples, like others in this study (see also subtypes A, C, F and I), which were collected in February 1994 from the AIDS clinic in Nicosia, Cyprus. PBMC DNA was PCR

amplified and cloned. Individual clones were sequenced. Patient 04 was a 24 year old man, asymptomatic with a CD4 count of 516, who had lived abroad, and had been seropositive for at least 4 years. Patient 11 was a 29 year old homosexual man, symptomatic with a CD4 count of 277, and seropositive for at least 7 years. Patient 21 was a 38 year old bisexual man, asymptomatic with a CD4 count of 743, who was infected in Greece and seropositive for at least 4 years. Patient 25 was a 34 year old heterosexual man, asymptomatic with a CD4 count of 650, who had lived in the U.S. and was seropositive for at least 6 years. Patient 27 was a 20 year old homosexual man, asymptomatic with a CD4 count of 709, infected in Cyprus and seropositive for at least 2 years. Patient 28 was a 39 year old homosexual woman, asymptomatic with a CD4 count of 430, infected in Cyprus and seropositive for at least 1 year. Patient 29 was a 49 year old heterosexual man, asymptomatic with a CD4 count of 420, infected in Cyprus and seropositive for at least 2 years. Patient 39 was a 37 year old heterosexual man, asymptomatic with a CD4 count of 470, had lived in Greece and was seropositive for at least 5 years. Patient 40 was a 29 year old heterosexual woman, symptomatic with a CD4 count of 92, who had lived in the U.S. and was seropositive for at least 6 years. Patient 43 was a 36 year old homosexual man, symptomatic with a CD4 count of 396, who had lived in the U.K. and Greece and had been seropositive for at least 6 years. Patient 45 was a 32 year old heterosexual man, asymptomatic with a CD4 count of 453, who was infected in Cyprus and seropostitive for at least 1 year. Patient 46 was a 32 year old heterosexual man, whose partner had died of AIDS, asymptomatic with a CD4 count of 107, had lived in the U.K. and was seropositive for at least 4 years. Patient 48 was a 22 year old hemophiliac man, asymptomatic with a CD4 count of 276, who had been seropositive for at least 11 years. Patient 50 was 32 year old homosexual man, whose partner had AIDS, asymptomatic with a CD4 count of 315, who had been seropositive for at least 1 year. GenBank accession numbers U28663, U28666 (04); U28664 (11); U28667–U28671 (21, 25, 27, 28, 29); U28675, U28676 (39, 40); U28678 (43); U28680–U28682 (45, 46, 48); U28684 (50).

21) **CZ.BTSPR:** This sequence is from a 32 year old homosexual man from Prague, Czech Republic. It was included as part of a study of HIV-1 sequences from the neighboring Slovak Republic. DNA from cocultured PBMCs was PCR amplified and the PCR product cloned. Eight clones were sequenced from each patient, although only one sequence is presented in the publication and in Genbank. It is not stated whether each single sequence is a consensus of the eight clones or a single clone. [Zachar et al.(1996b)]. GenBank accession number U53195.
22) **DE.D31:** This sequence is from isolate D31. [Kreutz et al.(1992)]. It has never been well described, it is only shown as HIV1-D31 in figure 3 of the paper. The complete genome has been sequenced. GenBank accession number U43096.
23) **DE.HAN:** This sequence is from an infectious clone from the German isolate DE.HAN-2. [Sauermann et al.(1990)]. GenBank accession number U43141.
24) **DE.Serocons:** This sequence is a consensus sequence from 7 hemophilia patients who all received the same lot of beta-propiolactone and UV-light inactivated clotting factor in Bonn or Goetingen, Germany, from November 1989 to March 1990. The virus and the patients have been extensively studied over time, since initial seroconversion. The sequences which were combined to create the consensus were from proviral DNA from cultured PBMCs, PCR amplified and cloned. [Kasper et al.(1994)]. GenBank accession numbers S76444 and S76446.
25) **DE1.ID#:** These 7 sequences are from a Neurology thesis by I. Weber of Rostock, Germany. They are from blood and CSF specimens from 6 patients. For each of the patients, phylogenetic analysis showed that blood and CSF sequences were more similar to each other than to other database sequences. For one patient, one of the two blood sequences was different enough from another blood and CSF pair to be included here. GenBank accession numbers Z78482, Z78485-Z78493.
26) **DE2.ID#:** These 4 sequences are from a set containing 15 Dutch homosexuals, 19 Dutch intravenous drug users, 2 German homosexuals, 2 German intravenous drug users, 5 Scottish homosexuals and six Scottish intravenous drug users. The sequences were used in a study of HIV-1 vpr, vpu, and env V3 regions and how they vary between risk groups. [Kuiken et al.(1996b)]. GenBank accession numbers Z68529, Z68530, Z68537 and Z68538.
27) **ES.ID#:** These 36 sequences are from 41 patients sampled in Madrid, Spain between 1985 and 1991. Proviral DNA was extracted from uncultured patient PBMCs and the C2V3 region was PCR amplified. The PCR products were directly sequenced. Two of the sequences reported in this set (D22-28 and D22-48) were 99.5% identical to the LAI (IIIB) lab strain of HIV-1 and are not included here. 3 other goups

of sequences had members that were greater than 98% identical to each other (R1, R2 and R3; THF13-2, THF12-24; S1, S4) and only one of each of them is presented here. [Quinones-Mateu et al.(1996)]. GenBank accession numbers U40533–U40552 and U45286–U45307.

28) **FR.J91:** This sequence is from one of the JBB clones from the French patient Bru. [Wain-Hobson et al.(1991)] and [Guo et al.(1991)]. GenBank accession numbers X57449–X57459 and X57461.

29) **FR.LAI:** This sequence is from the French isolate LAI (formerly BRU) which is also referred to as IIIB. [Wain-Hobson et al.(1985)]. Also see: [Alizon et al.(1986)], [Lukashov & Goudsmit(1995)] and [Wain-Hobson et al.(1991)]. GenBank accession numbers K02013, L23090–L23103, X01762, L48380–L48399, M64178–M64223, M64406–M64415 and M64768–M64775. Other sequences which are of this type include: PV22, K02083; MFA, M33943 [Stevenson et al.(1990)]; un-named, Z11530; BH8, K02011; BH10, M15654; TH4, L31963; MCK1, D86068; PM213, D86069; F12CG, Z11530; and HXB, K03455, M38432, M64775 and M14100. This isolate of HIV-1 has also been extensively studied in cases such as the infected lab worker. See for example [Jr et al.(1994)], [Pincus et al.(1994)] U12030–U12055. Recombinant virus pNL4-3, with envelope from LAI(BRU) and gag-pol from NY5 has also been studied: [Adachi et al.(1986)] GenBank accession number M19921, [Duensing et al.(1995)] GenBank accession number L42371 and [Salminen et al.(1995)] GenBank accession number U26942. Other GenBank entries with IIIB-LAI sequences can be found in the patented sequences section and in the cloning vector section (for example U19867 and A00647).

30) **FR1.ID#:** These 4 sequences are from Toulouse, France. In this study, 4 mother-infant pairs were followed during pregnancy and after birth. The inter- and intra-patient sequence similarities of this set of 308 sequences has been controversial, because some infant sequences were identical to sequences from other mothers. For purposes of this V3 section, only one sequence from each of the 4 infants is presented here. [Briant et al.(1995)], [Korber et al.(1995)] and [Learn et al.(1996)]. GenBank entries for all 308 sequences are found with accession numbers U24717-U24999 and U25001-U25025.

31) **GA.OYI:** This sequence is from the Gabonese isolate OYI (designated elsewhere as isolate 397), isolated from a healthy HIV-1 infected individual. GA.OYI appears to have been the first viral sequence from Africa that phylogenetically clustered with North American viruses. [Huet et al.(1989)]. GenBank accession number M26727.

32) **GB.AIT:** This sequence is from an individual at the time of seroconversion. Proviral DNA was extracted from PBMCs from a patient who was viraemic and had an equivocal HIV-1 antibody status, and the env V3 region was PCR amplified. The PCR products were cloned and the DNA sequence determined for 15 clones. These data showed that the V3 region contained only limited sequence heterogeneity with a major variant (shown here) accounting for 66% of the protein quasispecies present. [Ait-Khaled & Emery(1993)]. GenBank accession number S69598.

33) **GB.CAM1:** This sequence is from the British isolate CAM1. McIntosh A, and Karpas A, Thesis (1991), Cambridge University, England. GenBank accession numbers D10112, D00917.

34) **GB.Man** This sequence was PCR amplified from the 1959 "Manchester sailor" kidney tissue. The sequence of the complete genome is available and clusters with subtype B contemporary HIV-1 sequences. [Zhu & Ho(1995)]. GenBank accession number U23487.

35) **GB.V-ID#:** A set of six sequences from a study of hemophiliacs from Scotland who were originally thought to have been infected by the same batch of factor VIII. (ScV12 is a sequence from a hemophiliac from the U.S., included as a control). All are consensus sequences of multiple direct PCR sequences obtained from limiting dilution of PBMCs. The Scottish hemophiliacs were infected in 1984 and the PBMCs were obtained for analysis in 1989. Although the samples were potentially related, they were deemed sufficiently divergent in this region for inclusion in this set. [Simmonds et al.(1990)], and [Balfe et al.(1990)]. GenBank accession numbers M61327–M61346 and M61391–M61407. GenBank entries with accession numbers M84240–M84317 are more sequences from patient 82, taken over the period from 1984–1991. [Holmes et al.(1992)]. GenBank entries with accession numbers L13488–L13497 are also from these patients [Zhang et al.(1993)]

36) **GB1.CPHL1:** This is a consensus from the British isolate 93–08020, clones 1, 4, 7, 18, 19 and 43. It was referred to as 93–08020 in [Arnold et al.(1995c)] and was isolated from the patient referred to as CPHL1 in [Arnold et al.(1995a)]. CPHL1 is a surgeon and CPHL2 was a patient of his in 1986, approximately 7 years prior to sampling for this study. Because sequences from CPHL1 and CPHL2 are no more similar to each other than to sequences from the general population, transmission cannot

be concluded, and both sequences are included in this alignment. GenBank accession numbers U21100 (clone 1) and U23112–U23116 (clones 18, 19, 4, 43 and 7 respectively).

37) **GB2.CPHL2:** This is a consensus from the British isolate 93–17305, clones 3, 11, 18 and 25 [Arnold et al.(1995c)]. It was isolated from the patient referred to as CPHL2 in [Arnold et al.(1995a)]. GenBank accession numbers U23117–U23120 (clones 11, 18, 25 and 3 respectively).

38) **GB3.CPHL7:** This sequence is a sequence from the British isolate 94–24612, clone 13. [Arnold et al.(1995c)]. It was isolated from the patient referred to as CPHL7 in [Arnold et al.(1995a)]. GenBank accession number U23126. U23127 is a second clone from this same isolate. Sequences from three other patients epidemiologically linked to CPHL7 (CPHL6, accession numbers U23130–U23132; CPHL8, U23128–U23129; CPHL9, U23133–U23135) are not included in this alignment.

39) **GB4.ID#:** These sequences are from British isolates from St. Bartholomew's Hospital, London (M23470, M26864, M30156, M737677 and M737658) and Hammersmith Hospital, London (AC, JB and WB). Sequences were determined from cloned PCR products from PCR amplified DNA from either cocultured (M23470 and M26864) or uncultured (M30156, M737677, AC, JB and WB) patient PBMC proviral DNA. [Douglas et al.(1996)]. Complete envelope gp160 sequences were determined for at least one clone from each patient. Ugandan samples also sequenced in this report were subtypes D or D/A recombinant. GenBank accession numbers for London samples were U36859–U36864, U36869, U36870, U36872–U36880 and U36882.

40) **GB5.ID#:** These 11 sequences are from Scotland. They are part of a set containing 15 Dutch homosexuals, 19 Dutch intravenous drug users, 2 German homosexuals, 2 German intravenous drug users, 5 Scottish homosexuals and six Scottish intravenous drug users, from which regions of vpr, vpu and env were sequenced. The authors found consistent differnces in the sequences between the homosexuals and IV drug users. Only 34 of the 47 patients' sequences are reported in the publication. [Kuiken et al.(1996b)]. See also B_NL12 and B_DE2 sets.

41) **GM.GM6:** A sequence from Gambia, as yet unpublished. Bobkov et al. 1996 unpublished. GenBank accession number U33101. See also Gambian sequences of subtypes C and J.

42) **HT.RF:** This sequence is from the full-length lambda clone HAT-3, from Haitian isolate RF. RF is from a 28 year old Haitian male who had moved to the United States at age 25, in 1980. He had no history of IV drug use, homosexuality or blood transfusions. In October 1983 he had 20 lb weight loss, giardia with diarrhea, thrush, and diffuse lymphadenopathy. His CD4/CD8 ratio was 0.08. He died in December, 1983. Primary culture from a November 1983 blood sample was co-cultured on HUT-78 cells. [Jr et al.(1992)] [Starcich et al.(1986)] [Popovic et al.(1984)]. GenBank accession numbers M17451 and M12508.

43) **HT1.D-ID#:** These seven sequences are from Haitians, and are part of a set of sequences generated as part of the DAIDS variation program in the laboratories of Dr. Beatrice Hahn at the University of Alabama, and Dr. Marcia Kalish at the Centers for Disease Control, Atlanta, GA. Except for D2HA590, the full gp160 was sequenced from clones derived from expanded culture stocks. D2HA590 is a direct sequence from PCR amplified DNA from expanded culture. The sequence ID numbers are abbreviated, for example D2HA590 can be read as DAIDS sequence (D), isolated in 1992 (2), Haitian (HA), patient 301590 (590). Full length env for some of these have been expressed [Gao et al.(1996a)]. GenBank accession numbers: U08441–U08447, U04900. Both U08441 and U08442 are sequences from patient HT1.D1HA651 and are identical over the region of interest. GenBank accession numbers for additional clones derived from these patients: U04901–U04906.

44) **HT2.H-ID#:** These 25 sequences are from Haitians. All sequences were PCR amplified from the infected individuals PBMCs, and this set includes direct sequences of PCR amplification products, consensus sequences of multiple clones of PCR products plus one direct sequence, and single clones of PCR products. Full length env for some of these have been expressed [Gao et al.(1996a)]. These sequences were provided by the Centers for Disease Control, Atlanta, GA USA (Dr. Chin-Yih Ou), and John Hopkins University School of Hygiene and Public Health, Balimore, MD USA (Dr. Neal Halsy), and the Centers for Development and Health, Complexe Medico Sociale de la Cite Soleil, Port-au-Prince, Haiti (Dr. Reginald Boulos). GenBank accession numbers L07145–L07161, L07163–L07165, L07167–L07207, L07209–L07239, L07241–L07246, U08441–U08447.

45) **ID.1701:** This sequence is from a dried blood spot collected in 1992 from a male homosexual patient in Indonesia. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. [Cassol et al.(1996)]. GenBank accession number U53317.

46) **IN.IND9:** This sequence was from a heterosexually infected patient from New Delhi, India. DNA was isolated from cocultured PBMCs after one week of culture. PCR product was cloned and a single clone was sequenced. [Tripathy et al.(1996)]. GenBank accession number U31364. See also C_IN3.ID#.
47) **IN1.ID#:** These four sequences were isolated in Hyderabad, Andhra Pradesh, in southern India. The C2V3 region of env was amplified by nested priming from DNA from PBLs from fresh blood samples. Date of sampling and health status of HIV-1 infected individuals is unknown. [Baskar et al.(1994)]. GenBank accession numbers L29091–L29094.
48) **IT.ID#:** These two sequences are consensus sequences from 4 clones each. They were obtained from PCR-amplified proviral DNA from Langerhans cells from skin patches from a deceased AIDS victim in Italy [Sala et al.(1995)]. Small V1-V2 region sequences and V3-loop sequences from the same skin samples were published in [Sala et al.(1994)]. GenBank accession numbers U20670–U20677 used here. GenBank Accession numbers Z34376–Z34458, Z34470–Z34513 and Z34515 were V1-V2 and V3-loop sequences from the same patient.
49) **IT.PD:** This sequence is from the thymus of a 26 year old male with CDC stage III disease. Proviral DNA was PCR amplified from genomic DNA extracted from patient thymocytes, and the PCR product was directly sequenced. A V3 sequence from PBMC proviral DNA was also determined, as were gp41 sequences from both thymus and blood. [Calabro et al.(1995)]. GenBank accession numbers U09254–U09255.
50) **IT1.ID#:** These 10 sequences are from infants infected in utero. The sequences came from PCR amplified DNA of uncultured PBMCs, PCR amplified DNA of cultured PBMCs, or from RNA from serum collected at or shortly after delivery. [Scarlatti et al.(1993)]. GenBank accession numbers L08277–L08286. Sequences from the mothers of these infants are also available in entries with accession numbers L08287–L08372.
51) **IT2.ID#:** These 3 sequences are from unpublished sequences in GenBank entries with accession numbers X92424–X92426.
52) **JP.ETR:** A Japanese isolate from long-term cell culture with a truncated env gene, due to a point mutation of a CAG codon to a TAG stop codon. [Shimizu et al.(1992)]. GenBank accession numbers D12582, D01205–D01207, D12584 and D12571.
53) **JP.GUNA:** A Japanese 1989 isolate HIVGUN, infectious to T cells, was adapted to grow in fibroblast-like BT cells. A single amino acid change at the tip of the V3 loop was shown to be responsible for the change in tropism, GPGR to GSGR. [Takeuchi et al.(1991)]. GenBank accession number M59192.
54) **JP.JH32:** This is a sequence from a lambda clone of Japanese isolate JH3, which was isolated in 1986 from a 10 year old Japanese Hemophiliac. [Komiyama et al.(1989)]. GenBank accession number M21138.
55) **JP.KM03:** This sequence is from a 28 year old hemophilia B patient with CDC stage IV disease and T-cell count of 20, living in Japan. The authors [Hattori et al.(1991)] also sequenced the V3 region from 28 other Japanese individuals, but only the V3-loop amino acid sequence is available from the other patients. GenBank accession number S70936.
56) **JP.JNIH1M:** This sequence is from a Myanmarese (Burmese) individual living in Japan, obtained by direct sequencing of PCR-amplified proviral DNA from peripheral blood mononuclear cells. [Weniger et al.(1994)]. GenBank accession number L32084.
57) **KP.Kr121** This sequence is from an unpublished GenBank entry with accession number X93580.
58) **LT.LIT#:** These 5 sequences are from Lithuania. [Lukashov et al.(1995)]. GenBank accession numbers: L38417, LIT18A; L38412, LIT11A; L38419, LIT21A; L38416, LIT17A; L38420, LIT17A.
59) **MY1.ID#:** These two sequences are from Myanmarese (Burmese) individuals living in Myanmar, obtained by direct sequencing of PCR-amplified proviral DNA from peripheral blood mononuclear cells. [Weniger et al.(1994)]. Patient 02 is a male IV drug user, and 05 is a female prostitute, both were from Mandalay. GenBank accession numbers L32088, L32089.
60) **MY2.ID#:** These 11 sequences are from IV drug using prisoners in a prison in Kuala Lumpur, Malaysia. PCR products amplified from uncultured PBMCs were directly sequenced. [Brown et al.(1996)]. This report also included subtypes C and E in Malaysia. GenBank accession numbers U65538–U65548.
61) **MY3.ID#:** These 5 sequences are from dried blood spots collected in 1992 from a male STD patient (1782), a female prostitute (1739), and 3 IV drug users (1748, 1755 and 1763) in Myanmar. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. [Cassol et al.(1996)]. GenBank accession numbers U53304–U53308.

62) **NL.114C** This consensus sequence represents sequences generated from PCR amplified plasma RNA from one of three infants in a Dutch mother/infant study. Samples were collected from the infant at birth, at 6 weeks and at 9 months of age. Samples were also collected from the mother before birth, at birth and after birth. Mother sequences are not included in this consensus. [Mulder-Kampinga et al.(1993)]. [Mulder-Kampinga et al.(1995)]. Infant 114 is from GenBank accession numbers L21111–L21153. Mother 114 sequences are from GenBank accession numbers L21028–L21110. Infant 127 sequences are from GenBank accession numbers Z47817–Z47832. Mother 127 sequences are from GenBank accession numbers Z47833–Z47880. Gag gene sequences from mother/child pairs are also available in Genbank accession numbers Z47903–Z47911; Z47912–Z47928; Z47929–Z47935; Z47936–Z47950. The second mother/child pair was also from the Netherlands, see G_NL.127C. The third mother/infant pair in this study was from Rwanda, see A_RW.564C.

63) **NL.168:** This is a consensus sequence of 3 clones after culturing in PBMC. The isolate was originally from an AIDS patient in Amsterdam. [Wrin et al.(1995)]. GenBank accession numbers U15030–U15032. A V3-loop (105 bp) segment from the original isolate has been previously reported. [Fouchier et al.(1992)]. GenBank accession number L06694.

64) **NL.wolfscon:** This is a consensus sequence from 6 patients infected with the same unit of blood in Amsterdam. [et al(1992)]. GenBank accession numbers U04530–U04537.

65) **NL.X1con:** This is a consensus sequence of 10 clones from a recipient in a donor-recipient study. Sequences from donor Y and recipient X2 are also part of this study, but are not included here. [Cornelissen et al.(1995)]. GenBank accession numbers Z47505–Z47514 are from X1. Other new sequences analyzed in this paper include Z47411–Z47540. Sequences M91828–M91838 (donor H and recipient O, referred to as patients A14 and A13, respectively in [Wolfs et al.(1992)] see B_NL1.A13) were also re-analyzed in this study.

66) **NL1.ID#:** These eight sequences are part of a study of presumed donor-recipient pairs from an HIV-1 transmission study conducted in the Netherlands. If pairs were extremely close or identical, only the recipient is included here. Recipient samples were from the first sample to be antibody positive, and are numbers 1,3,5,7,9, and 13. These sequences are consensus sequences of multiple clones from PCR amplified serum RNA. [Wolfs et al.(1992)]. Recipient A1 was also studied as patient H1 in [Kuiken et al.(1993)] and so is not included here (see B_NL4.H1). GenBank accession numbers M91819–M91827, M91829, M91831–M91832, M91839, M91857–M91870, M91872, M91874, M91881–M91884, M91891, M91893, 91895–M91908, M91910, M91911–M91926. Number 13, and the donor were also analyzed in [Cornelissen et al.(1995)].

67) **NL2.ID#:** These two sequences are part of a Dutch study of mutations occurring over a five year period (starting in 1985) in two patients. Serum RNA was PCR amplified and multiple clones were sequenced. The consensus for each patient is shown. [Wolfs et al.(1991)] and [Zwart et al.(1992)]. GenBank accession numbers M74591–M74684.

68) **NL3.NET-ID#:** These six consensus sequences from the Netherlands are samples from AIDS patients, using serum samples to generate PCR clones from viral RNA for sequencing. [Zwart et al.(1993)]. GenBank accession numbers L01282–L01297.

69) **NL4.ID#:** These 74 sequences represent a study of early seroconverters from different times with different risk factors for transmission during the AIDS epidemic in the Netherlands. The year the sample was taken is indicated in the last part of the sequence name. The risk group of the individual from whom the virus is derived is indicated in the first letter of the sequence name (I, B and H for IVDUs, hemophiliacs, and homosexuals, respectively). Viral genomic RNA from sera was PCR amplified and amplification product was directly sequenced. [Kuiken et al.(1993)]. GenBank accession numbers Z29219–Z29225, Z29256–Z29258, and Z29262–Z29325.

70) **NL5.ID#:** These 18 consensus sequences are from a study of patients with, and without, AIDS dementia complex (ADC) in the Netherlands. Not all patients were from the Netherlands. Samples were collected between 1986 and 1992. Viral genomic RNA from sera and/or cerebral spinal fluid was reverse transcribed and PCR amplified and clones were sequenced. [Kuiken et al.(1995)] GenBank accession numbers Z37531–Z37534, Z37734–Z37963, Z37970–Z37971

71) **NL6.ID#:** 16 is a consensus sequences of four sequences used in a study of HIV-1 envelope-mediated syncytium formation. The consensus represents four clones from one patient, two clones of the consensus are SI and two are NSI. 320 is a single SI clone. The sequences were derived from PCR amplified DNA

from provirus cultured in MN cells. [Andeweg et al.(1992)]. GenBank accession numbers L08655–L08662.

72) **NL7.ID#:** These two consensus sequences are from sets of sequences (GenBank accession numbers U05797, U13240, U13241, U13243–U13247 for consensus 537 and U13242, U13248–U13252 for 1058) used in a study on the dynamics of HIV sequence changes in vivo and the utility of heteroduplex analysis. Both sequences were derived from PCR amplified PBMC DNA. Consensus 537 represents a set of sequences from a Dutch patient with a relatively stable CD4+ cell count at 62 months post-seroconversion. Consensus 1058 represents sequences from another Dutch patient whose CD4+ cell count at 73 months post-seroconversion was declining faster than 537's. [Delwart et al.(1994)]. See also US10.ID#.

73) **NL8.672** This is a sequence from a patient early in infection, before, or around the time of seroconversion. Three other patients studied in this paper (537, 1058 and 594), had previously been reported. See B_NL4.594, B_NL7.1058con and B_NL7.537con. [Shpaer et al.(1994)] and [Delwart et al.(1995)]. GenBank accession numbers U23651–U23663, U23667, U23670. See also US11.ID#.

74) **NL9.h#:** These sequences are from a cohort of homosexual men living in Amsterdam who seroconverted between 1985 and 1989. The sequences are from direct sequencing of PCR products after RT-PCR from serum RNA. Samples for h1, h139, h491, h1140 and h1234 were obtained at seroconversion. Samples for h138 and h1136 were obtained 12 months and 29 months after seroconversion respectively. The sequence for h320 was obtained from proviral DNA, 2 months after seroconversion. [Zwart et al.(1994a)] and [Zwart et al.(1994b)]. GenBank accession numbers L25884 and U05786–U05808. More sequences from this same cohort of men were published in [Kuiken et al.(1996b)] and [Kuiken et al.(1996a)] GenBank accession numbers Z67875–Z67876, Z67885–Z67939, Z67941–Z67960, Z68015–Z68089, Z68109–Z68110. Some vpr, vpu and other region sequences are available from some of these patients as well. Some of the GenBank entries for this set appear to be duplicates of sequences reported in other studies, for example patient H537 entries U05797 and

75) **NL10.ID#:** These 13 sequences are from recent immigrants to The Netherlands from various countries. The fist two letters of the ID# represent the two letter country code for the previous residence of the patient. The next two numerals represent the year of isolation. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced. [Lukashov et al.(1996)]. GenBank accession numbers L76878, SR903853; L76897, SR9304737; L76879, SR911515; L76874, UM893272; L76912, UM9403860; L76888, SR925752; L76890, SR926969; L76894, GQ9301341; L76882, MA913670; L76886, SR923572; L76873, BR891413; L76910, UM9403051; L76885, GF921953.

76) **NL11.ID#:** These 2 sequences are consensus sequences of 81 clones (patient N) and 105 clones (patient F) from serum, sigmoid tissue and fecal matter from each patient. All sequences from patient N were more similar to other sequences from patient N than to any other sequence in the database. Likewise all sequences from patient F were most similar to other patient F sequences. van der Hoek et al. In Press (1996). GenBank accession numbers Z76463–Z76648.

77) **NL12.ID#:** These sequences are part of a set containing 15 Dutch homosexuals, 19 Dutch intravenous drug users, 2 German homosexuals, 2 German intravenous drug users, 5 Scottish homosexuals and six Scottish intravenous drug users, from which regions of vpr, vpu and env were sequenced. The authors found consistent differnces in the sequences between the homosexuals and IV drug users. Only 34 of the 47 patients' sequences are reported in the publication. [Kuiken et al.(1996b)]. See also B_GB5 and B_DE2 sets. Some of the patients in this study have been previously studied. For example, entries with accession numbers Z68061 and U05787 are both from the same patient.

78) **NO1.ID#:** These 36 sequences are from unpublished Norwegian sequences with Genbank accession numbers X92902–X92912, X92915, X92916, X92919–X92941. Four subtype C sequences were also part of this set (X92913, X92914, X92917, X92918).

79) **PR1.D-ID#:** These four sequences are from Puerto Rico, and were generated as part of the DAIDS variation program in the laboratory of Dr. Marcia Kalish at the the Centers for Disease Control, Atlanta, GA. The C2V3 region was directly sequenced from PCR amplification of DNA from viral culture. The sequence ID numbers are abbreviated; for example D2PR732 can be read as DAIDS sequence (D), isolated in 1992 (2), Puerto Rico (PR), patient 301732 (732). A full description of these sequences can be found in the April 1994 supplement to the HIV database, part III. GenBank accession numbers U04926–U04929.

80) **PY.ID#:** Ten sequences from 10 patients living in Asuncion, Paraguay. All 10 were male patients with symptoms of AIDS. Virus was propagated in tissue culture for an unstated length of time prior to

harvesting proviral DNA for PCR and sequencing. PCR products were directly sequenced. PY.3614, PY.3615 and PY.12837 were syncytium-inducing and the other 7 were not. The tenth sequence, PY.3614p was PCR amplified directly from patient PBMCs with no culturing. The significant differences between the cultured sequence from this patient (PY.3614c) and the direct sequence, indicates that the virus that grew out in culture was a minority of the virus present in PBMC. The sequence of PY.3614c is not included in this alignment. [Cabello et al.(1995)]. GenBank accession numbers U28949–U28959.

81) **RU.RUS#:** These 2 sequences are from [Lukashov et al.(1995)]. GenBank accession numbers: L38405, RUS3A; L38407, RUS4A.
82) **RU2.ID#:** These 15 sequences are from homosexuals living in St. Petersburg, Russia. Malykh et al unpublished (1995). GenBank accession numbers U40283–U40285 and U40319–U40330.
83) **SE.pt11s113:** This sequence is from one (patient 11 sample 113, collected in 1988) of a set of 13 sample from 9 epidemiologically linked individuals. The index case (patient 1) was a Swedish male who is believed to have contracted HIV while visiting Haiti in 1980. Six Swedish females were infected (patients 2,4,5,7,8 and 11) by patient 1. Two males (patients 6 and 10) were then infected by these females, and two HIV-infected children (patients 3 and 9) were born to the women. Sequences from each patient were determined by PCR amplification from uncultured PBMCs and direct sequencing. Heterogeneous sites were indicated with IUPAC codes. Extensive phylogenetic analysis was done to determine which methods accurately reconstructed the true phylogeny. [T. Leitner et al.(1996)]. GenBank accession numbers U68496–U68521.
84) **SE1.ID#:** Seven sequences that are consensus sequences of blood and CSF samples taken from each patient. The CDC disease stage class for the patients are as follows: II pts 930, 2815; III pt 931; IV-E pt 2951; IV-A pt 1032; and IV-C2 pts 1433, 1866. [Keys et al.(1993)]. GenBank accession numbers Z23178–Z23181, Z23185–Z23187, Z23192–Z23195, Z23200–Z23219, Z23224–Z23227, Z23232–Z23235, and Z23240–Z23255.
85) **SE2.ID#:** These five sequences are from patients in Goetebotg, Sweden each of whom had recently seroconverted at the time of sampling in 1985-1990. Sequences from the sexual partners of all five patients who were believed to have transmitted the virus to these recipients were also determined, but are not shown here due to the great similarity to the recipient sequences (94% to 99% identity). Three of the recipients (R1, R3 and R4) were homosexual, and the other two were heterosexual. [Furuta et al.(1994)]. GenBank accession numbers U10929–U10950.
86) **SK1.ID#:** These nine sequences are from homosexual men living in the inner city of Bratislava, in the Slovak Republic. Patients 11, 12, 15 and 23 were classified as CDC stage A1 and were not taking any medication. Patients 18 and 20 were CDC stage B2 and were taking AZT. Patients 9 and 28 were CDC stage C3 and were taking AZT. No information is available for patient 22. Two other patients (10 and 51) were included in this study, but their sequences are not shown here because after their sequences proved to be similar to sequences from patients 9 and 28, respectively, epidemiological investigation indicated that they (9 and 10; 28 and 51) had been sexual partners. DNA from cocultured PBMCs was PCR amplified and the PCR product cloned. Eight clones were sequenced from each patient, although only one sequence is presented in the publication and in Genbank. It is not stated whether the single sequence is a consensus of the eight clones or a single clone.[Zachar et al.(1996b)]. GenBank accession numbers U53192–U53194, U53196–U53203.
87) **TH.93TH067:** This sequence is from Thailand. It is one of several complete env gene sequences obtained for the World Health Organization. HMA subtyping as well as sequence-determined subtyping was done on each one. Sequences were PCR amplified from cocultured PBMCs. Two to three clones from each isolate were sequenced. [Penny et al.(1996)]. GenBank accession numbers U39258 and U39259.
88) **TH.T8174:** This sequence comes from a study of the genetic heterogeneity and epidemiological distribution of HIV1 in Thailand. The host was an intravainous drug user and the sequence was obtained from PCR amplified PBMC DNA. [Ou et al.(1993)]. GenBank accession number L19238. See also E_TH.T8178.
89) **TH1.ID#:** These ten sequences are from individuals from Thailand. PCR-direct, peripheral blood PBMC DNA. [Ou et al.(1992b)] and [Ou et al.(1993)]. (Published erratum appears in Lancet **342**:250 (1993).) GenBank accession numbers L07442, L07449–L07456 and L07460.
90) **TH2.ID#::** The TB132 sequence is from a set of isolates from HIV seropositive individuals from Thailand. PCR, PBMC co-culture, DNA. Full env sequence is available. [McCutchan et al.(1992)]. Please note:

the TB132 locus name in the database corresponds to the McCutchan et al. "BK132" isolate. GenBank accession number L03697. The CM237 sequence is from PBMC proviral DNA. [Mascola et al.(1994)]. GenBank accession number L14570. See also B_US14

91) **TH3.W2TH-ID#:** 2 sequences from Thailand from asymptomatic individuals. Consensus, PCR-clones, cell-culture DNA and RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April 1994 supplement to the *Human Retroviruses and AIDS 1993* database. [De Wolf et al.(1994)]; [Osmanov et al.(1994)]; [Gao et al.(1994a)]. GenBank accession numbers U08715–U08718, U08801–U08802, and U08783–U08784.

92) **TH4.ID#:** These twelve sequences are B subtype sequences from Thailand. Ten were genetically most similar to HIV-1 found in the Americas and Europe; these sequences were derived from people infected prior to 1988 (diagnosed in 1986 or 1987). The other two (N762 and N763) were designated B' and were isolated from people with more recent infections, 1988 and 1992. The sequences were obtained from PCR amplified PBMC DNA. The naming of the sequences submitted to GenBank does not correspond with the naming of the sequences in the paper. [Kalish et al.(1994)]. GenBank accession numbers for the entire set of thirteen sequences studied in this publication: U15576–U15588.

93) **TH5.ID#:** These three sequences are B subtype sequences from Thailand. Two individuals believed to be dually infected with subtypes B and E were analyzed. It is not clear from the paper or the GenBank entries, which sequences came from individual 1 and which from 2. [Artenstein et al.(1995)]. Genbank accession numbers U21471, U21473, U21475. See also E_TH6.ID#.

94) **TH6.ID#:** These 4 entries are from 1992 dried blood spot samples from Thailand. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. Samples came from 3 previously identified HIV seropositive IV drug users (58, 11, 15) and a homosexual (86). Other samples from the same region were all subtype E (see E_TH8.#). [Cassol et al.(1996)]. GenBank accession numbers U53310, U53311, U53314 and U53315.

95) **TH7.ID#:** These 70 sequences are from IV drug users in Bangkok, Thailand, who were undergoing methadone treatment at 14 treatment clinics. Blood samples were collected between January and April, 1994. Uncultured PBMC DNA from each patient was PCR amplified, and the PCR product was directly sequenced (except for patient 108, in which PCR product was cloned and 2 clones were sequenced, one shown here). Of the 84 patients sampled, 69 were Thai B, one (091) was typical subtype B, and 14 were subtype E. [Kalish et al.(1995)]. GenBank accession numbers U22543–U22547, U22549–U22552, U22554–U22556, U22558–U22560, U22562–U22566, U22568–U22574, U22576–U22603, U22605–U22608, U22610, U22613–U22616, U22618–U22623 and U22626. See also E_TH9.

96) **TT.QZ5489:** This sequence is from Trinidad. W. Blattner et al. Unpublished 1995. Genbank accession number U32396.

97) **TW1.ID#:** These 16 sequences are from healthy HIV-1 carriers or AIDS patients from Taiwan [Chang et al.(1996)]. Three subtype B sequences in this set were greater than 97% identical to the HXB2/LAI lab strain (see B_FR.LAI) of HIV-1, and are not included here (TW83, U73049; TW271, U73059; and TW335, U73061). The manuscript reports that 123 of 143 sequences from Taiwan were subtype B, but only 27 of the 143 sequences were submitted to the sequence databases. Other subtypes found in Taiwan in this study were E (17 cases), C (1 case), F (1 case) and G (1 case). GenBank accession numbers for B subtype are U73045–U73054, U73056, U73057, U73059, U73061 and U73063–U73069.

98) **US.ACP1:** This virus was cultured from a seronegative man with Kaposi's sarcoma. (See: [Ho et al.(1989)]). ACP1 was the sequenced after one passage in PBMCs. GenBank accession number M80660. The sequence AC-H9 (GenBank M80661) was also derived from this patient. [Ashkenazi et al.(1991)].

99) **US.ADA:** This sequence is from the monocytropic U.S. isolate ADA. [Westervelt et al.(1991)]. GenBank accession number M60472.

100) **US.ALA1:** This sequence is from an infectious clone of the 1985 U.S. isolate AL-1, taken from a patient with AIDS. Buckler-White, A. et al., Unpublished (1988). GenBank accession number M38430.

101) **US.BAL1:** This sequence is from the macrophage tropic U.S. isolate BAL, harvested from lung alveolar tissue. Reitz M, et al., Unpublished (1990). GenBank accession numbers M68893, M68894. GenBank entry M63929 is 98% identical to this sequence and is also derived from the BAL isolate [Hwang et al.(1991)].

102) **US.BCSG3:** This is a fragment of a full genomic sequence from the provirus SG3, cloned as a single proviral unit. This clone replicates more efficiently in chimpanzeee than in human lymphocytes, and is extremely cytopathic in immortalized human T-cell lines. [Ghosh et al.(1993)]. GenBank accession number L02317.

103) **US.BRVA:** This sequence is from U.S. isolate BRVA, which was taken from the brain tissue of a AIDS patient with neurological disorders. [Anand et al.(1989)]. GenBank accession number M21098.

104) **US.BWB:** This consensus sequence is from sequences derived from PCR amplified PBMC DNA from brain tissue. [Monken et al.(1995)]. GenBank accession numbers L17088–L17126.

105) **US.CDC42:** This sequence is from an infectious clone of the U.S. isolate CDC-451. [Desai et al.(1986)]. GenBank accession number M13137.

106) **US.Diaz:** This sequence is one of a set of 124 closely related sequences. All 124 sequences came from patients who had been infected by a common source. Unpublished (1995) Diaz,R., Zhang,L., Busch,M., Mosley,J. and Mayer,A. Genbank accession numbers for all 124 sequences are U29433–U29437, U29956–U30054, and U43035–U43054. Gag sequences from 12 clones are also available in GenBank entries U31573–U31584.

107) **US.JFL:** This sequence is from a non-infectious clone from the monocytropic U.S. isolate JFL. [McNearney et al.(1990)]. GenBank accession number M31451. Other sequences from HIV-1 isolates epidemiologically linked to this isolate can be found in GenBank entries with accession numbers L06256–L06273. [McNearney et al.(1993)].

108) **US.JM:** This sequence, along with B_US.WM, came from viral isolates after short term culture in PBMCs, PCR amplification, and cloning of PCR products. Both are from asymptomatic, seropositive individuals. [Ashkenazi et al.(1991)]. GenBank accession number M80662.

109) **US.JRCSF:** This sequence is from an infectious clone of the 1986 U.S. isolate JRCSF, derived from from the CSF of a patient who died with Kaposi's sarcoma and severe AIDS encephalopathy. The infectious clone JRFL was isolated from the brain of the same patient. [O'Brien et al.(1990)]. GenBank accession numbers U63632, M38429, U45960. Also see: [Pang et al.(1991)], [Klasse et al.(1996)].

110) **US.MN:** This sequence is from 1984 U.S. isolate MN, taken from a 6 year old male pediatric AIDS patient from the Newark, New Jersey area in 1987. His mother was an IV drug user who died of pneumonia in 1982. His father was also HIV seropositive. A complete genome of isolate MN is found in M17449. Other sequences from this patient from the 1984 blood sample and from a 1987 sample taken shortly before death (U72495) are available also. [Jr et al.(1992)] [Gurgo et al.(1988)]. GenBank accession number M17449. See also L48364–L48379 [Lukashov & Goudsmit(1995)].

111) **US.NY5CG:** This sequence is from the 1984 U.S. isolate NY5. [Willey et al.(1986)]. GenBank accession number M38431. See also GenBank accession number K03346. A recombinant between NY5 and LAI has also been extensively studied, see B_FR.LAI entry.

112) **US.P896:** This sequence represents a molecular clone from an primary isolate derived from a Jamaican man who immigrated to Philadelphia 15 years earlier. At the time of viral isolation, he had no antiviral therapy, but was an AIDS patient with $<$ 10 CD4 cells per mm3. The infectious molecular clone from which this sequence was derived is both macrophage-tropic and extremely cytopathic in lymphocytes. [Collman et al.(1992)] and [Kim et al.(1995)]. GenBank accession numbers M96155, U39362.

113) **US.RJS:** This is a consensus of six biologically characterized clones from patient RJS, isolate 4. The HIV-1 infected individual had been infected for five years at the time of isolation. Complete env sequence is available. [Daniels et al.(1991)] and [Fisher et al.(1988)]. GenBank accession numbers M37491 and M37573–M37577.

114) **US.SB(A-C):** These three sequences are from 1988 U.S. isolates taken from a woman, her daughter and her sexual partner. The three viruses are epidemiologically linked, however the amino acids sequences appeared sufficiently divergent in this region to merit the inclusion of all three samples. Sequences were directly sequenced from PCR amplification products after the virus was briefly cultured. [Burger et al.(1991)]. GenBank accession numbers M77228–M77230.

115) **US.SC:** This sequence is from the 1984 U.S. isolate SC, from an AIDS patient. [Gurgo et al.(1988)]. GenBank accession number M17450.

116) **US.SF128:** This clone was isolated from the spinal cord tissue of a patient with dementia, after coculture with PBMCs, and infects macrophages but not T-cells [Liu et al.(1990)]. GenBank accession numbers M95292 and M38673. The ability to infect macrophages vs. HUT 78 cells was mapped to the region

between a StuI site in env and a XhoI site in nef, by replacing this region in SF2 with the same region from SF128.

117) **US.SF162:** This sequence is from an infectious clone from the U.S. isolate SF162, cultured from the cerebrospinal fluid of a patient with toxoplasmosis. [Cheng-Mayer et al.(1990)]. GenBank accession numbers M38428, M65024.

118) **US.SF2:** This sequence is from an infectious clone from the U.S. isolate ARV-2. ARV-2/SF2 was isolated from a patient with oral candidiasis after co-culture with mitogen-stimulated PBMCs in 1984 [Levy et al.(1984)]. [Sanchez-Pescador et al.(1985)]. GenBank accession numbers K02007 and I07977. HIVSF13 (GenBank accession number L07422) is a more infectious virus taken from the same patient five months later, when he had developed Kaposi's sarcoma and Pneumocystis carinii pneumonia [Cheng-Mayer et al.(1991)]. SF2 and SF13 are 98% identical to one another.

119) **US.SF33:** This sequence is from an infectious clone from the 1984 U.S. isolate SF33. [York-Higgins et al.(1990)]. GenBank accession number M38427.

120) **US.TN-ID#:** These eight sequences are from asymptomatic individuals identified after donating blood in Memphis, Tennessee, USA. [Slobod et al.(1994)]. GenBank accession numbers U09140–U09175.

121) **US.twinABcon:** This sequence is a consensus of a set of 27 env sequences from a pair of heterozygotic perinatally HIV-1 infected twins who were observed during their first 2 years of life. Twin A remained asymptomatic through her first 2 years while twin B developed AIDS at six months and died at 22 months of age. Patient PBMCs were cocultured with donor PBMCs, prior to DNA extraction and PCR amplification. Approximately 500 copies of HIV proviral DNA per PCR reaction were used. PCR products were cloned prior to sequencing. Viral phenotypes from both infants, and all time points were also assessed. All were found to be non-syncytium inducing, but they differed in their ability to infect primary macrophages. Overall, it seems that the production of neutralizing antibodies by the healthy twin was the most important clinical factor. [Hutto et al.(1996)]. GenBank accession numbers U47562-U47588 for the envelope sequences and U47589-U47613 for tat sequences from the same blood samples.

122) **US.WM:** This sequence, along with B_US.JM, came from viral isolates after short term culture in PBMCs, PCR amplification, and cloning of PCR products. Both are from asymptomatic, seropositive individuals. [Ashkenazi et al.(1991)]. GenBank accession number M80663.

123) **US.WMJ22:** This sequence is from the isolate WMJ22, isolated from a person of Haitian descent living in the U.S. Virus was cocultured on an immortalized T-cell line. [Hahn et al.(1986)] and [Starcich et al.(1986)]. GenBank accession numbers M12507, K03457.

124) **US.WR27:** This sequence is from a complete genome. It represents first complete PCR-derived sequence of a U.S. clinical isolate of genotype B expanded only in primary PBMC. This provirus harbors a uniquely truncated V3 loop. [Salminen et al.(1995)]. GenBank accession number U26546.

125) **US.YU:** This is a consensus sequence of eight lambda phage clones and 12 PCR amplified clones derived from the uncultured brain tissue of a patient with AIDS dementia complex. A macrophage tropic clone (YU-2) is almost identical to the consensus sequence of YU in this region, with only a single amino acid change (K to N) 10 amino acids from the carboxy-terminal end of the sequence. [Li et al.(1991)]. Complete genomic sequences are available for two of the HIV1YU clones, along with biological characterizations of four of the HIV1YU clones: [Li et al.(1992)]. The GenBank accession numbers for the YU-10 and YU-2 complete genomes are M93259 and M93258, respectively. GenBank accession numbers for the other clones are M89972–M89984.

126) **US1.HC-ID#:** These are forty 1990–1991 U.S. samples, from the study of the dentist who was thought to have been the source of HIV-1 infection of six of his patients. Only the dentist's viral sequence and the Florida control sequences are shown here; the six epidemiologically and genetically linked patients are excluded from this alignment because their viral sequences were very similar to the dentist's. All sequences were PCR amplified from patient PBMCs. Most are direct sequences from the amplification products, although some are consensus sequences of multiple clones of PCR products and one direct sequence. [Ou et al.(1992a)]. GenBank accession numbers for the 75 sequences in this set: M90847–M90853, M90881–M90886, M90894–M90900, M90907–M90912, M90914–M90956, M90958–M90964, M92100–M92133, L22590–L22606 and U06872-U06919. See also [Korber & Myers(1992)], [Crandall(1995)], [Smith & Waterman(1992)], [DeBry et al.(1993)], [Palca(1992b)], [Palca(1992a)], [Abele & DeBry(1992)], [Hillis & Huelsenbeck(1994)], [Ciesielski et al.(1992)] and [Ciesielski et al.(1994)].

127) **US2.D-ID#:** These 15 sequences are from the USA, and are part of a set of sequences generated as part of the DAIDS variation program in the laboratory of Dr. Marcia Kalish at the the Centers for Disease Control, Atlanta, GA. The C2V3 region was directly sequenced from PCR amplification products of DNA from viral culture. The sequence ID numbers are abbreviated, for example D2US711 can be read as DAIDS sequence (D), isolated in 1992 (2), United States (US), patient 301711 (711). A full description of these sequences can be found in the April 1994 supplement to the HIV database, part III. Full length envelope genes from some of these clones have been expressed, [Gao et al.(1996a)]. GenBank accession numbers U04907–U04915, U04918 and U04921–U04925, U08448–U08452.

128) **US3.D-ID#:** These four sequences are from the US, and are part of a set of sequences generated as part of the DAIDS variation program in the laboratory of Dr. Beatrice Hahn at the University of Alabama. gp160 sequences of clones from expanded culture stocks are available. The sequence ID numbers are abbreviated, for example D2US711 can be read as DAIDS sequence (D), isolated in 1992 (2), United States (US), patient 301711 (711). A full description of these sequences can be found in the April 1994 supplement to the HIV database, part III. GenBank accession numbers U08448–U08449 and U08451–U08452. GenBank accession numbers for additional clones from these patients: U04916–U04917 and U04919–U04920,

129) **US4.ZhuPt-ID#:** Sequences from five primary seroconverters. Consensus, PCR-clones, peripheral blood PBMC DNA. Although sequences were also available from two of the donors, only sequences from five recipients are shown here. [Zhu et al.(1993)]. GenBank accession numbers L21224–L21324, L21372–L21373, L21376–L21377, L21380–L21381, L21384–L21389, L21393–L21394, L21397–L21398, L21400–L21401, L21404, L21405, L21408, L21410–L21411, L21414, L21417–L21418, L21426–L21427, L21430–L21432, L21434–L21437, L21440–L21442, L21445, L21447, L21449–L21450, L21453–L21454, L21457, L21459, L21462–L21463, L21519–L21520, L21522–L21523, L21525–L21526, L21528–L21533, L21535–L21536, L21538–L21539, L21541, L21543, L21545–L21546, L21548, L21550–L21551, L21553–L21554, L21556, L21558–L21559, L21561–L21569, L21571–L21578, L21580, L21582, L21584, L21586, L21588 and L21590.

130) **US5.pt-ID#:** A consensus of PCR-clones, peripheral blood DNA. These two sequences were part of a study of blood sequences compared to brain sequences from six individuals. [Korber et al.(1994)]. GenBank accession numbers U05360–U05568.

131) **US6.-ID#:** Each sequence is a consensus sequence of several cloned PCR products from PBMC proviral DNA from an individual infant. These sequences were part of a study of mother-infant transmission. Infant blood samples were taken from 1 week (infant 7) to 34 months (infant 4) post-partum. Dates of sample collection were: Infant 1, 10/25/91; Infants 2–4, 10/31/91; Infant 5, 2/6/92; Infant 6, 8/30/93; Infant 7 5/13/93. Maternal sequences were also reported as part of this set, but are not used in building these consensus sequences. [Ahmad et al.(1995)]. GenBank accession numbers U16390–U16652.

132) **US7.-ID#:** Each sequence is a consensus sequence of several cloned PCR products from PBMC proviral DNA from an individual patient. These sequences were part of a study of early samples (1984-1986) from the San Francisco region. One of the samples (552-3) was HIV negative but was determined to be contaminated with blood from another (565-3) so the six samples from 552-3 and 565-3 are made into one consensus here (565). Another sample (552-5) from patient 552 was not contaminated, and it is presented as B_US11.552, because a larger region was reported in that publication. [Sabino et al.(1994b)] GenBank accession numbers L20371–L20380. More V3 and tat sequences from these individuals are discussed in [Sabino et al.(1994a)]. GenBank accession numbers U00243–U00399 and U01513–U01529.

133) **US8.-ID#:** These sequences were from a study of three recipients of contaminated blood. Recipient 1 (R1) and recipient 2 (R2) each received blood from different donors. A third recipient, not presented here, received blood from both donors. All three recipients were neonates. R1 received erythrocytes from donor 1 on 19 October, 1984 at the age of 3.5 weeks. For R1 the sequence of one of the two clones (2E) is presented here. R2 received erythrocytes from donor 2 on 24 September, 1984 at the age of 2 months. For R2 a consensus of 6 clones is presented here. Blood samples for this study were drawn in March 1986. R1 had slow weight gain, and R2 had lymphadenopathy at time of sample collection. [Diaz et al.(1995)]. GenBank accession numbers U11188 = R1–2E, U11189 not used; U11203, U11196, U11199, U11192–U11194 = R2 six clones. The tat and envelope V4-V5 regions of clones from these same individuals are also available in U11173–U11178, U11180, U11205–U11209.

134) **US9.-ID#:** These are consensus sequences for samples taken over a range of time from four different subjects. Blood samples for S1 were drawn in Nov '85, Jul '87, Jan '88 and May '89. Blood samples for S2 were drawn in May '85, Apr '87 and Oct '87. Blood samples for S3 were drawn in Jun '87 and Dec '87. Blood samples for S4 were drawn in Jan '85, Jan '89 and Jun '89. S2, S3 and S4 had decreasing CD4 counts during the study period. S1 had fluctuating CD4 counts. [McNearney et al.(1992)]. GenBank accession numbers L03430–L03453 and L23575–L23588 = S1; L03454–L03477 and L23618–L23633 = S2; L03478–L03490 and L23589–L23600 = S3; L03491–L03515 and L23601–L23617 = S4.

135) **US10.ID#:** These three consensus sequences are from sets of sequences used in a study on the dynamics of HIV sequence changes in vivo and the utility of heteroduplex analysis. All sequences were derived from PCR amplified PBMC DNA. The MA145 consensus represents sequences (GenBank accession numbers U00821, U00822, U00831–U00839) taken from an asymptomatic male from Massachusetts over a period of 4.5 years starting April 1989. Patient MA, from the US, was infected in 1984 or 1985, and had been experiencing neurological disorders prior to 1989 [Kusumi et al.(1992)]. The SFBU and SFPE consensuses represent sequences (GenBank accession numbers U13373–U13380 and U13381–U13388 respectively) taken from two patients with AIDS from San Francisco. [Delwart et al.(1994)]. Other sequences from patient MA can be found in GenBank entries with accession numbers U00804–U00822, U00831– U00850, U00873–U00888, M79342–M79354 and M90025–M90046. Other SFBU sequences can be found in U13240–U13252. See also NL7.ID#.

136) **US11.ID#:** These four consensus and seven individual sequences came from patients early in infection, before, or around the time of seroconversion. Sequences for 306, 419, 349, and 074 are consensus sequences. [Shpaer et al.(1994)] and [Delwart et al.(1995)]. GenBank accession numbers U23664–U23666, U23668, U23669, U23671–U23708, L20381. See also B_NL8.ID#.

137) **US12.ID#:** These sequenses were used in an investigation into the transmission of HIV-1 from one child (CHA), who had received zidovudine, to another child (CHB), who harbored a zidovudine-resistant strain. The presence of the zidovudine-resistant strain in child 1 and 2, and the lack of such a strain in child 2's mother was used to show that child 2 was infected by child 1 and not by child 2's mother. LC sequences are from children used as local controls. All sequences were derived from PCR amplified PBMC DNA. [Fitzgibbon et al.(1993)]. GenBank accession numbers L12751–L12756, L19695, L19697, S66942. L12756 is listed as "isolate 100" in GenBank, but seems to be the "group B consensus sequence" used for phylogenetic analysis.

138) **US13.ID#:** These three consensus sequences are from three IV drug users in Florida. Proviral DNA sequences were obtained from blood, cerebrospinal fluid and dorsal root ganglia from each of the three individuals. Sequences for V1-V5 of env, were PCR amplified, cloned and sequenced. [Shapshak et al.(1995)]. [Xin et al.(1995)]. The sequence for patient 149 is a consensus of all 24 clones in GenBank accession numbers U16094–U16117. The sequence for 141 is a consensus of GenBank entries labelled as being from patient 141 with the exception of: R5D, R6D, R7D, R2D and R4D, which were similar to IIIB strains of HIV-1; and R1R, R3R, R7R, R8R and R9R, which were similar to samples from patient 144. The sequence for 144 is a consensus of GenBank entries labelled as being from patient 144 with the exception of: R3R, R6R, R9R, R12R, R13R and L1D, which were similar to IIIB strains of HIV-1; and C3D, C4D, C7D, C8D and C10D, which were similar to samples from patient 141. While infection of each individual with multiple strains of HIV (including one very similar to the IIIB lab strain) is a possible explanation of these findings, we are only including one consensus sequence from each patient for this alignment. The authors are currently (1996) resequencing new samples from these patients. GenBank accession numbers for patients 141 and 144 are U16032–U16093.

139) **US14.ID#:** These four sequences are from DNA from PBMC. [Mascola et al.(1994)]. GenBank accession numbers L14573–L14576. See also B_TH2, E_TH2.

140) **US15.ID#:** These six consensus sequences are from a study of infants. Blood samples were collected from six infants over time. [Strunnikova et al.(1995)]. GenBank accession numbers U22682–U22810, U22834 and U22835.

141) **US16.ID#:** These two consensus sequences were used in a study of the impact of sequence variation on the distribution and seroreactivity of linear antigenic epitopes. Both sets of sequences were from PCR amplified DNA from peripheral blood leukocytes. Patient ARTC1 was an asymptomatic individual from New York and ARTC3 was an AIDS patient from New York. [Pestano et al.(1995)]. GenBank accession numbers U11586–U11594. See also A_UG1.964, C_UG1.45, and D_UG7.ID#.

142) **US17.ID#:** CB7 is a consensus sequence for 12 clones from two different samples (six clones each) from the same patient, collected in 1988 and 1990, plus two more clones as yet unpublished. The patient seroconverted in 1985. The patient did not receive any antiviral therapy until 1992. The patients CD4 count was 1035 in 1988 and 807 in 1990. The patient's PBMS were cocultured with donor PBMC for an unspecified length of time before cultured DNA was isolated and PCR amplified. Individual clones of PCR product were then sequenced. [Wang et al.(1995)]. GenBank accession numbers U16324–U16335 and U19706–U19711. The other 8 sequences are individual clones from 8 different patients, all from the same clinic in Boston, MA. Individual CB7 was again included in this study, [Wang et al.(1996a)] [Wang et al.(1996b)]. GenBank accession numbers U60152–U60162 and U27658–U27669.

143) **US18.ID#:** These 22 sequence are from a study of a HIV-1 infected dentist and many of his HIV-1 infected patients. In this case, no evidence was found for dentist to patient, patient to dentist, or patient to patient transmission. [Jaffe et al.(1994)]. GenBank accession numbers U11454–U11490. Unfortunately the dentist sequences are not available from GenBank.

144) **US19.ID#:** These two sequences are from a comparison of a mother who transmitted HIV to her infant and a mother who did not. One of the sequences (nontransmitter-283 accession number U07839) was 99.4% identical to the LAI strain of HIV-1 and is not included here. Sequences from the transmitter and nontransmitter were highly similar with the exception of sequence nontransmitter-217 (U07836). Thus a consensus of the similar sequences is included here as VA1-consensus, and the nontransmitter-217 outlyer is presented seperately as VA2. [Ayyavoo et al.(1996)]. GenBank accession numbers U07833–U07841 and U07891–U07917.

145) **US20.ID#:** These three sequences are from a study of mother-infant transmission. Only child consensus sequences are used here, but mother and child clonal sequences are reported in [Roth et al.(1996)]. Proviral DNA was PCR amplified from uncultured patient PBMCs. Cloned PCR products were sequenced. Child 1 was 27 months old, child 2 was 5 months old and child 3 was 21 months old when blood was drawn for sequencing. Sequences, including those from the mothers, are in GenBank entries with accession numbers U47745–U47807.

146) **US21.ID#:** These six sequences are from six different patients, all men from the Chicago, Illinois, USA MACS cohort. In the study, P1 & P2 were rapid progressors, P3 & P4 normal progressors, and P5 & P6 non-progressors. Their estimated years of seroconversion are as follows: P1, 1985; P2, 1985; P3, 1986; P4, 1986; P5, 1985; and P6, 1984. A total of 292 sequences of the C2-V5 region of envelope were completed. In the database entries the sample numbers are of the form Px.y-z, where x stands for the patient number, y is the number of months after the estimated date of seroconversion, and z is the clone number. Each of the six sequences presented here is a consensus of all clones from that patient. [Wolynsky et al.(1996)]. GenBank accession numbers U35895–U36185.

147) **US22.ID#:** These 12 sequences are from a study of variability of tat and env genes. Lorenzo et al (1996) Unpublished. GenBank accession numbers U57104–U57216. Tat sequences from these clones are also available (U57217–U57304).

148) **US23.ID#:** These 3 sequences are from a study of variability of vif and env genes. [Sova et al.(1995)]. Uncultured or short-term cultured PBMC proviral DNA was PCR amplified, and several clones from each PCR reaction were sequenced. Only one clone sequence from each patient is presented here. GenBank accession numbers U50615–U50628. Vif sequences from these patients are also available (U41055, U41056, U41179–U41182 and U42229–U42282).

149) **US24.ID#:** These 5 sequences are from unpublished GenBank entries by Schwartz et al. GenBank accession numbers U49518–U49640, U51311–U51326 and U45860–U45876.

150) **VE.ID#:** These 8 sequences are from 8 individuals in Venezuela. Patient PBMCs were cocultured with donor PBMCs. Proviral DNA was harvested PCR-amplified. PCR products were directly sequenced. Nearly complete env gp120 sequences were determined, as well as pol gene sequences. [Quinones-Mateu et al.(1995)]. GenBank accession numbers U16764–U16778, even numbers are env, odd numbers are pol.

151) **VN1.HCM9:** This sequence is from South Vietnam. [Menu et al.(1996)]. The sequence is from Ho Chi Minh city, from a woman infected by her HIV seropositive sexual partner who was thought to have been infected while traveling in Europe. Three other sequences in this study were found to be subtype E. Genbank accession number U29209.

152) **ZA.ID#:** These 7 sequences are from 7 individuals in South Africa. ZA504 was from a 33 year old white male homosexual with AIDS and the virus was syncytium-inducing (SI). ZA508 was from a 32

year old white male bisexual with ARC and the virus was NSI. ZA509 was from a 30 year old white male homosexual with AIDS and the virus was SI. ZA524 was from a 49 year old white male bisexual with AIDS and the virus was NSI. ZA510 was from a 29 year old white male heterosexual with ARC and the virus was SI. ZA512 was from a 26 year old white male homosexual with ARC and the virus was SI. ZA513 was from a 3 year old black male blood transfusion recipient with AIDS and the virus was SI. All samples were collected at the Tygerberg Hospital in the Western Cape region of South Africa between 1984 and 1992. DNA was harvested from cocultured PBMCs and the env gene was PCR amplified and cloned into pBSKS+ for sequencing. Each sequence is from a single cloned PCR product. [Engelbrecht et al.(1995)]. GenBank accession numbers L48063–L48066, L48069, L48071 and L48073. GenBank entries U33770 and U33774–U33779 are shorter env gene fragments from these same clones.

C Subtype

At this time there are viral sequences from 119 HIV-1 infected individuals associated with HIV-1 subtype C. The C subtype consensus sequence (C_CONSENSUS_96) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published and/or have been made available for printing in the database by their authors.

1) **BR.W2BR025:** This sequence is part of a gp160 sequence from an asymptomatic individual from Brazil, sampled in 1992. A clone was derived from an expanded viral culture, expressed and sequenced. This sequence was provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences from this patient can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. Relevant papers are: [De Wolf et al.(1994)]; [Osmanov et al.(1994)]; [Gao et al.(1994a)]. GenBank accession number U15121. Entries with accession numbers U08720, U08785, U09126, U09132 and U09133 are also from W2BR025.
2) **BR1.HSP203** Although this sequence is listed as unpublished in the database, it seems to be an extension of work published in [Morgado et al.(1994)] and [Sabino et al.(1994c)]. It is from San Paulo, Brazil. GenBank accession number U31585.
3) **BR2.91BR015:** This sequence is from Brazil. It is one of several complete env gene sequences obtained for the World Health Organization. HMA subtyping as well as sequence-determined subtyping was done on each one. Sequences were PCR amplified from cocultured PBMCs. Two to three clones from each isolate were sequenced. 91BR015 is similar to another WHO sample; 92BR025. [Penny et al.(1996)]. GenBank accession numbers U39234 and U39238.
4) **BU1.ID#:** These sequences are from Burundi. They are part of several complete env gene sequences obtained for the World Health Organization. HMA subtyping as well as sequence-determined subtyping was done on each one. Sequences were PCR amplified from cocultured PBMCs. Two to three clones from each isolate were sequenced. 91BU009 groups with subtype D in a neighbor- joining tree of the V3 region and is presented as an undetermined subtype in the section of the compendium. [Penny et al.(1996)]. GenBank accession numbers U39252 and U39233, 91BU001; U39248 and U39237, 91BU002; U39239 and U39242, 91BU003; U39241 and U39243, 91BU004; U39240 and U39257, 91BU005; U39244 and U39246, 91BU006; U39245, U39247 and U39249, 91BU007; U39250 and U39251, 91BU008; U39253 and U39254, 91BU009.
5) **BY1.ID#:** These 3 sequences are from Byelorussia. [Lukashov et al.(1995)]. GenBank accession numbers: L38410, BLR9A; L38409, BLR8A; L38408, BLR5A.
6) **CF1.15166:** A single sequence from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990–1991. It is a consensus from PCR clones from cultured proviral DNA. [Murphy et al.(1993)]. GenBank accession number L11525. Other sequences from this study were subtypes A and E.
7) **CY.HO02:** This is a sequence from a 51 year old woman whose husband had died of AIDS. She was born and lived in Zambia, before moving to Cyprus. She was asymptomatic, with a CD4 count of 200, and she had been seropositive for at least 6 years. This sample, like others in this study (see also subtypes A, B, F and I) was collected in February 1994 from the AIDS clinic in Nicosia, Cyprus. DNA was extracted from patient PBMCs and PCR amplified. After a second round of PCR, products were cloned and sequenced. Two clones from patient 02 were sequenced. [Kostrikis et al.(1995)]. Genbank accession numbers U28321 and U28661.
8) **DJ1.DJ-ID#:** These two sequences from Djibouti were from a a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession numbers L22940 and L23065.
9) **ET2.ID#:** These nine sequences of Ethiopian isolates were part of a cohort considered to be heterosexually infected. PBMC DNA was PCR amplified and directly sequenced. [Salminen et al.(1996)]. For patient 2220, who had slim disease and AIDS, almost full length HIV-1 genome was PCR amplified from PBMC DNA and cloned. Proviral DNA from others in the cohort was PCR amplified and directly sequenced. Other regions of the genome are available for these isolates as well. [Salminen et al.(1996)]. GenBank accession numbers U45481–U45502 (V3 region), U15060–U15066 and U45503–U45504 (LTR NF-kB/NRE regions), M64001–M64009 (gag p7 region), M64015–M64018 (env gp41 region) and U46016 (C2220 complete genome).

10) **GA1.ID#:** These 2 sequences are from Gabon. G134 is from a 1988 or 1989 sample from a patient with AIDS living in Franceville, Gabon. LBV105 is from a 1988 sample from an asymptomatic individual sampled from the general population of Libreville. Method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced. [Delaporte et al.(1996)]. GenBank Accession numbers X90912, G134; X90913, LBV105. See also subtypes A, D, F, G and O sequences from this same study.
11) **GB1.00513:** This sequence is from the British isolate 93–00513. [Arnold et al.(1995c)]. GenBank accession number U21099.
12) **GM.GM3:** A sequence from Gambia, as yet unpublished. Bobkov et al. 1996 unpublished. GenBank accession number U33098. See also Gambian sequences of subtypes B and J.
13) **IN1.D-ID#:** These four sequences are from samples from high risk patients in India, PCR clones, DNA, PBMC culture. [Dietrich et al.(1993)]. GenBank accession numbers L07651 and L07653–L07655; X65638–X65640 and X68406.
14) **IN2.D-ID#:** These five sequences are from samples from high risk patients in India, primarily stage I. They were nested PCR amplified from DNA obtained from uncultured PBMC from patients serologically defined as HIV-1/HIV-2 mixed infections. [Grez et al.(1994)]. GenBank accession numbers U07098 and U07100–U07103.
15) **IN3.IN#:** These 8 sequences sequence were isolated from Pune and New Delhi, India. All 8 sequences were from heterosexually infected patients from New Delhi, or Pune, India. DNA was isolated from cocultured PBMCs after one week of culture. PCR product was cloned and a single clone was sequenced. [Tripathy et al.(1996)]. GenBank accession numbers U29179, U29694–U29698, U31362 and U31363. See also B_IN.IND9.
16) **IN4.ID#:** These 24 sequences are from 1992 dried blood spot samples from Vellore near Madras, in Tamil Nadu state in southern India. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. Samples came from previously identified HIV seropositive STD patients (1, 3, 5, 6, 7, 11, 12, 13, 19, 20, 23, 29, 33), spouses of infected men (2, 4, 10, 16, 22, 26, 27), female prostitutes (36, 37) or bisexual men (8, 32). Two homosexual men from the same region had subtype A HIV-1 (see A_IN1.9 and 14). [Cassol et al.(1996)]. GenBank accession numbers U53278–U53285, U53287–U53290, U53292–U53303.
17) **KE.NA113:** This sequence was derived from a patient who was part of a May-June 1992 study of pregnant women from the Pumwani Maternity Hospital in Nairobi, Kenya. Viral RNA was concentrated from patient serum just prior to delivery, and the envelope C2-V3 region was amplified by RT-PCR. The PCR product was cloned and 20 clones from the patient were sequenced. Seven other patients from this study had viral subtypes A and D. [Zachar et al.(1996a)]. Genbank accession number U33762.
18) **MW.SM750:** This sequence is from cloned PCR amplified cocultured PBMC DNA. SM750 was a black male sampled in the gold mines in Malawi in 1989. [Becker et al.(1995)]. GenBank accession number U06719.
19) **MW1.ID#:** These 13 sequences are from pregnant women with risk factors from Malawi. PCR-direct, peripheral blood DNA. [Orloff et al.(1993)]. GenBank accesssion numbers L07427–L07441, L15721–L15735.
20) **MW2.D-ID#:** These two sequences are from individuals from Malawi, generated as part of the DAIDS variation program in the laboratory of Dr. Beatrice Hahn at the University of Alabama. The C2V3 was excised from full gp160 sequences, derived from clones from expanded culture stocks. The sequence ID numbers are abbreviated, for example D3MA959 can be read as DAIDS sequence (D), isolated in 1993 (3), Malawi (MA), patient 301959 (959). GenBank accession numbers U08453–U08454.
21) **MY.ID#:** These 2 sequences are from IV drug using prisoners in a prison in Kuala Lumpur, Malaysia. PCR products amplified from uncultured PBMCs were directly sequenced. Both of these prisoners had received medical treatment in India which included blood transfusion and organ transplants, and it is likely that they were infected in India. [Brown et al.(1996)]. This report also included subtypes B and E in Malaysia. GenBank accession numbers U65549–U65550.
22) **NL1.ID#:** These 2 sequences are from a Dutch woman whose partner was a recent immigrant to The Netherlands from Zaire, and from a recent immigrant to The Netherlands from either Zambia or Zaire. The fist two letters of the ID# represent the two letter country code for the previous residence of the patient (UN = unknown). The next two numerals represent the year of isolation. Viral RNA was prepared

from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced. The sequence from the Dutch/Zaire patient is not unequivocably a subtype C sequence, it may be subtype A or unclassifiable. [Lukashov et al.(1996)]. GenBank accession numbers L76908, NL9402418; L76898, UN9305091.

23) **NO1.ID#:** These 4 sequences are from unpublished Norwegian sequences with Genbank accession numbers X92913, X92914, X92917 and X92918. Thirty-six subtype B sequences were also part of this set (see B_NO1.ID#:).
24) **RU.ID#:** These 4 sequences are from Russia. [Lukashov et al.(1995)]. GenBank accession numbers: L38418, RUS20A; L38404, RUS2A; L38406, RUS1A; L38414, RUS13A.
25) **RU1.ID#:** This sequence is from Russia. Bobkov et al. 1996 Unpublished. GenBank accession number U33109.
26) **RW1.ID#:** These two sequences are consensus sequences of many clones from mothers infected in Rwanda. Mulder-Kampinga et al., Unpublished (1996). Mother 566 was apparently dually infected with HIV-1 subtypes A and C. See also A_RW4.ID#. GenBank accession numbers for mother 566 are Z76183, Z76188–Z76197, Z76284–Z76293, Z76295–Z76299, Z76301, Z76303–Z76311, Z76322–Z76342 and Z76725–Z76732; for mother 134, Z76039–Z76042 and Z76044.
27) **SN.SE364:** A Senegalese sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession number L22944.
28) **SO.1574:** This sequence is a consensus sequence of blood and CSF samples taken from the Somalian patient 1574, CDC classification II. [Keys et al.(1993)]. GenBank accession numbers Z23188, Z23190–Z23191, and Z23228–Z23231.
29) **SO.SM145:** A Somalian sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession number L22946.
30) **TW1.252** This sequence is one of a set of sequences from Taiwan. Other sequences in the set were subtypes B, E, F or G. [Chang et al.(1996)]. GenBank accession number U73055.
31) **UG.45:** A single sequence used in a study of the impact of sequence variation on the distrinution and seroreactivity of linear antigenic epitopes. The sequence was derived from PCR amplified DNA from peripheral blood leukocytes. The patient was an asymptomatic individual from Uganda. [Pestano et al.(1995)]. GenBank accession number U11597. See also A_UG1.964, B_US17.ID#, and D_UG7.ID#.
32) **UG.UG268:** A Ugandan sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession number L22948.
33) **ZA.NOF:** This sequence is from a South African individual who was part of a study of HIV-1 strains in India. This sequence was found to be closer to the Indian sequences, than are other isolates from Africa. PCR amplified DNA from PBMC cultures were sequenced. [Dietrich et al.(1993)] and [Becker et al.(1995)]. GenBank accession numbers L07426 and U06716. See also C_IN3.ID#, B_IN.IND9.
34) **ZA1.ID#:** These 2 sequences are from 2 individuals in South Africa. ZA514 was from a 59 year old mixed-race male heterosexual with AIDS and the virus was NSI. ZA517 was from a 33 year old mixed-race male heterosexual with ARC and the virus was NSI. The samples were collected at the Tygerberg Hospital in the Western Cape region of South Africa between 1984 and 1992. DNA was harvested from cocultured PBMCs and the env gene was PCR amplified and cloned into pBSKS+ for sequencing. Each sequence is from a single cloned PCR product. [Engelbrecht et al.(1995)]. GenBank accession numbers L48067–L48068. GenBank entries U33781 and U33782 are shorter env gene fragments from these same clones. See also B_ZA and D_ZA sequences from this same study.
35) **ZA2.ID:** These sequences are from clones from PCR amplified cocultured PBMC DNA. Dlu was a black male sampled at the Tygerberg Hospital in Cape Town, South Africa in 1990. Gom was a black male sampled at the Tygerberg Hospital in Cape Town, South Africa in 1990. BooyD was a mixed-race female mother sampled at the Tygerberg Hospital in Cape Town, South Africa in 1990. A short sequence from

a seropositive child of BooyD is presented in Genbank entry U07015. [Becker et al.(1995)]. GenBank accession numbers U07237, U06717 and U06718.

36) **ZM1.ZAM-ID#:** Two sequences from Zambia, from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession numbers L22954 and L22956.

37) **ZW.2647:** This sequence is a consensus sequence taken from blood and CSF samples taken from Zimbabwe patient 2647, CDC classification II. [Keys et al.(1993)]. GenBank accession numbers Z23196–Z23199 and Z23236–Z23239.

D Subtype

At this time there are viral sequences from 107 HIV-1 infected individuals associated with HIV-1 subtype D. The D subtype consensus sequence (D_CONSENSUS_96) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published and/or have been made available for printing in the database by their authors.

1) **BI.BU009con:** This is a consensus sequence from 8 clones from a 36 year old female with CDC stage IV AIDS and pulmonary tuberculosis, from Bujumbura, Burundi. PBMCs from the patient were cocultured with donor PBMCs for 2 weeks, at which time large syncytia formed and the culture became antigen positive. The env region was then PCR amplified by nested PCR and cloned. [Ranjbar et al.(1995)] and [Ranjbar(1995)]. GenBank accession numbers L35452–L35459.
2) **CF.4020:** This sequence was the only D subtype from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990–1991. It is a consensus from cloned PCR products, derived from cultrued proviral DNA. [Murphy et al.(1993)]. A full gp120 sequence from this isolate was kindly provided prior to publication by Dr. MP Kieny of Transgene, Strasbourg, France. It is a part of a set of 14 HIV-1 gp120 isolates from Bangui, Central African Republic. The gp120 proteins have been expressed in a hybrid gp160 vaccinia virus expression system [Richalet-Secordel et al.(1994)]. The year of isolation and health status of individuals from which the viruses were isolated were not provided. Viruses were isolated in the CAR by C Mathiot and B You (Pasteur Inst., Bangui), grown by F Barre-Sinoussi and A Deslandres (Pasteur Inst., Paris), and cloned and sequenced by D Schmitt and MP Kieny. GenBank accession numbers L11472–L11473, U43138.
3) **CI.CI-13:** A single D subtype sequence from a set of 13 isolates from individuals from Abidjan, Cote d'Ivoire. CI-13 was symptomatic, and serologically dually reactive for HIV-1 and HIV-2. The C2V3 region is part of a 900 bp fragment that was sequenced for each individual. Samples were collected between May 1990 and Sept. 1991. Virus was cultured with donor PBMCs, nested PCR amplified, 3–4 clones were sequenced, and the consensus of those clones is presented here. [Janssens et al.(1994a)]. GenBank accession numbers X72028–X72029.
4) **GA1.ID#:** These 2 sequences are from 1988 or 1989 samples from patients with AIDS living in Franceville Gabon. Method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced. [Delaporte et al.(1996)]. GenBank Accession numbers X90919, G109; X90920 G141. See also subtypes A, C, F, G and O sequences from this same study.
5) **GB1.CPHL4:** This sequence is a consensus from the British isolate 93–43424, clones 2, 6 and 35. It was referred to as 93–43424 in [Arnold et al.(1995c)] and as CPHL4 in [Arnold et al.(1995a)]. GenBank accession numbers U21098 (clone 35) and U23121–U23122 (clones 2 and 6 respectively). CPHL4 is a female who is believed to have contracted the virus from CPHL5 through heterosexual contact. Sequences from CPHL5 are not included in this alignment due to this epidemiological relationship.
6) **KE.NA116:** This sequence was derived from a patient who was part of a May-June 1992 study of pregnant women from the Pumwani Maternity Hospital in Nairobi, Kenya. Viral RNA was concentrated from patient serum just prior to delivery, and the envelope C2-V3 region was amplified by RT-PCR. The PCR product was cloned and 20 clones from the patient were sequenced. Seven other patients from this study had viral subtypes A and C. [Zachar et al.(1996a)]. Genbank accession number U33765.
7) **KE1.KEN-ID#:** These three patients were part of a 1990–1992 cohort study of maternal risk factors in mother to child transmission, including 22 pregnant women and an infant from Kenya. The C2V3 region was sequenced. [Janssens et al.(1994d)]. GenBank accession numbers for sequences from the entire set of 23 patients studied in this publication: U12984–U13006.
8) **NL.A11:** This sequence is from a Dutch study of presumed HIV-1 donor-recipient pairs. This sequence is from a recipient at the time of seroconversion; the donor was a Zairean woman living in the Netherlands (patient A12 GenBank Accession numbers M91840–M91848). The sequences from both donor and recipient were extremely similar, so only the recipient (patient A11) is shown here. This sequence is a consensus sequences of multiple clones from PCR amplified serum RNA. [Wolfs et al.(1992)]. GenBank accession numbers M91849–M91856.
9) **NL2.ID#:** These 7 sequences are from recent immigrants to The Netherlands from various countries. The first two letters of the ID# represent the two letter country code for the previous residence of the patient. The next two numerals represent the year of isolation. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were

directly sequenced. [Lukashov et al.(1996)]. GenBank accession numbers L76900, UG9307184; L76907, ZR9402261; L76892, ZR929193; L76904, UG9401525; L76895, AO9302187; L76872, ZR891183; L76876, ZR901100.

10) **RU.RUS14A** This sequence is from [Lukashov et al.(1995)]. GenBank accession number: L38415.
11) **SN.SE365:** A Senegalese sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession number L22945.
12) **TZ1.TAN-ID#:** These ten sequences are from a set of 14 Tanzanian samples from symptomatic individuals, using serum samples taken in 1988 to generate PCR clones from viral RNA for sequencing. [Zwart et al.(1993)]. GenBank accession numbers L01298–L01339.
13) **TZ2.ID#:** These eight sequences were from patients at a clinic in Dar es Salaam, Tanzania. The individuals from which the virus was cultured showed clinical signs of AIDS, and the year of viral isolation was 1988. Viral cDNA was PCR amplified from donor PBMC, and one cloned PCR product per donor was sequenced. [Siwka et al.(1994)]. GenBank accession numbers U12406, U12407, U12410–U12415.
14) **TZ3.ID#:** These 4 sequences are from the Mara region of rural northwest Tanzania [Robbins et al.(1996)]. Subtype A was also found in this study. GenBank accession numbers U61875 and U61879–U61881.
15) **UG.U44342:** This sequence is from a Ugandan. Consensus of PCR-clones, peripheral blood DNA. Intact env sequences are available from this sample. [Bruce et al.(1993)]. GenBank accession numbers M98408–M98416.
16) **UG.UG23:** This sequence is from blood collected from the Mulago Teaching Hospital in Kampala, Uganda. Viral RNA was harvested after 10-14 days of coculture with donor PBMCs and reverse-transcribed with AMV-RT. The env V3 region was PCR amplifed and cloned. This sequence is from an individual clone. [Atkin et al.(1993)], [Pestano et al.(1993)]. GenBank accession number M98504.
17) **UG1.W2UG-ID#:** Twelve sequences from asymptomatic individuals from Uganda in 1992. Each sequence is a consensus from cloned PCR products derived from cell-cultured proviral DNA or culture supernatant RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. [De Wolf et al.(1994)]; [Osmanov et al.(1994)]; [Gao et al.(1994b)]. GenBank accession numbers U08721–U08741, U08786–U08787, U08803–U08809 and U08821–U08824.
18) **UG2.ID#:** These twelve sequences are from 1986–1987 Ugandan samples. Each sequence is a consensus from cloned PCR products derived from uncultured proviral DNA harvested directly from patient PBMCs. [Oram et al.(1991)]. No GenBank entries exist for these sequences.
19) **UG3.ID#:** These 11 sequences are part of a set of sequences derived from 22 Ugandans who were attending an AIDS clinic. Blood samples were obtained in 1990. Each sequence is a consensus from cloned PCR products derived from uncultured proviral DNA harvested directly from patient PBMCs. [Albert et al.(1992)]. GenBank accession numbers L00614–L00618, L00733–L00737, M98894–M98899, M98901, M98906–M98907, M98911–M98913, M98918, M98920–M98923, M98929–M98937, M98942–M98945 and M98967–M98975.
20) **UG4.ID#:** One to 4 clones of each these Ugandan isolates were sequenced, but only one clone is shown here. [Douglas et al.(1996)]. Other Ugandan isolates sequenced in this study were subtype D/A recombinant. London subtype B clones were also reported. Complete envelope gp160 sequences were reported for all isolates. GenBank accession numbers U36867, U36868, U36871, U36884–U36887.
21) **UG5.UG-ID#:** Three Ugandan sequences from a set of HIV-1 viral isolates from Africa. Health status of the individuals from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession numbers L22947 and L22949–L22950.
22) **UG6.ID#:** Three Ugandan sequences from a set of HIV-1 viral isolates from Africa. All three individuals from which the virus was cultured had AIDS, and the year of viral isolation was 1987. Viruses were cultured with HUT-78 cells for an unspecified length of time. The V3 region of env (gp160) was amplified, cloned and sequenced. [von A. et al.(1995)]. GenBank accession numbers U15005, U15006 and U15007.

23) **UG7.ID#:** These sequences were used in a study of the impact of sequence variation on the distribution and seroreactivity of linear antigenic epitopes. Both sets of sequences were from PCR amplified DNA from peripheral blood leukocytes. All patients were asyptomatic individuals reporting for regular blood drawing at the Nakasero blood transfusion service, Kampala, Uganda. [Pestano et al.(1995)]. GenBank accession numbers U11595, U11596, and U11598. See also A_UG1.964, C_UG1.45, and D_UG7.ID#.
24) **UG8.#:** These 2 sequences are from Gulu, northern Uganda. They are from direct sequence of PCR product amplified from uncultured PBMCs. Blood samples were drawn from 217 pregnant women attending the clinic in Gulu, northern Uganda. Ages ranged from 17 to 37 years. The 29 seropositive women (13.4%) were all asymptomatic. [Buonaguro et al.(1995)]. Genbank accession numbers U44881 and U44884. Eight subtype A sequences were also found in this study (see A_UG5).
25) **US.AMK:** This sequence comes from a student living in Alabama, who moved from Zaire to the US in 1988. Virus was isolated from the patients PBMCs; this isolate was PCR amplified, and amplification products from both gag (U08192) and env were subcloned and sequenced. His CD4 count was < 5 cells/mm3, and he was symptomatic at the time of viral isolation. [Gao et al.(1994b)] and [Gao et al.(1996a)]. AMK is also known as 93ZR001. GenBank accession numbers U08193, U27419.
26) **ZA.ID#:** These 5 sequences are from 5 individuals in South Africa. ZA500 was from a 41 year old white male homosexual with ARC and the virus was non-syncytium-inducing (NSI). ZA501 was from a 24 year old white male bisexual with AIDS and the virus was SI. ZA505 was from a 36 year old white male homosexual with AIDS and the virus was SI. ZA506 was from a 33 year old white male homosexual with AIDS and the virus was SI. ZA507 was from a 37 year old white male homosexual with AIDS and the virus was SI. All samples were collected at the Tygerberg Hospital in the Western Cape region of South Africa between 1984 and 1992. DNA was harvested from cocultured PBMCs and the env gene was PCR amplified and cloned into pBSKS+ for sequencing. Each sequence is from a single cloned PCR product. [Engelbrecht et al.(1995)]. GenBank accession numbers L47608, L48061, L48062, L48070 and L48072. GenBank entries U33769, U33771–U33773 and U33780 are shorter env gene fragments from these same clones. See also B_ZA and C_ZA1 sequences from this same publication.
27) **ZR.ELI:** This sequence is from the Zairean isolate ELI. [Alizon et al.(1986)] and [Goodenow et al.(1989)]. The complete genomic sequence and an infectious clone are available. GenBank accession numbers M27949, K03454 and X04414.
28) **ZR.JY1:** This sequence is from Zairean isolate Z-84, clone JY1. [Yourno et al.(1988)]. GenBank accession number J03653.
29) **ZR.MAL:** This sequence is from a non-infectious clone of the Zairean isolate MAL. [Alizon et al.(1986)]. The complete genomic sequence and an infectious clone from the isolate MAL are available. MAL is known to be recombinant between subtypes A and D. GenBank accession numbers K03456 and X04415.
30) **ZR.NDK:** This sequence is from an infectious clone of the Zairean isolate NDK. The molecular clone is highly cytopathic in vitro. [Spire et al.(1989)]. The complete genomic sequence is available. GenBank accession number M27323.
31) **ZR.Z2Z6:** This sequence is from an infectious clone of Zairean isolate Z2. Theodore T, and Buckler-White A, unpublished. The complete genomic sequence is available. GenBank accession number M22639. See also [Srinivasan et al.(1987)]. GenBank entry with accession numbers K03458 and M16322, which is from the same isolate.
32) **ZR1.ID#:** These four sequences are part of a set of 14 A and D sequences from women from Zaire. 8 were healthy, 4 showed minor signs of illness, and 2 had AIDS. Sequences were determined by directly sequencing PCR products derived from uncultured proviral DNA harvested directly from patient PBMCs. [Potts et al.(1993a)]. GenBank accession numbers L19623, L19627, L19631 and L19635.

E Subtype

At this time there are viral sequences from 124 HIV-1 infected individuals associated with HIV-1 subtype E. The E subtype consensus sequence (E_CONSENSUS_96) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published and/or have been made available for printing in the database by their authors.

1) **CF.90CR402** 90cr402, previously named CAR-E 4002, was obtained from a man from Bangui, Central African Republic, who had lymphadeopathy, diarrhea, severe weight loss and recurrent respiratory infections. He was infected through heterosexual contact, but the year of infection is unknown. The virus was first adapted to growth in chimpanzee cells, expanded in chimpanzee cells, and then re-expanded in human PBMCs before lamba cloning and sequencing. [Gao et al.(1996b)]. The complete genome is found with GenBank accession number U51188.
2) **CF.4039:** This is a subtype E virus from the Central African Republic. Schmitt,D. et al. unpublished (1995). GenBank accession number U43112.
3) **CF1.ID#:** These eight sequences are from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990–1991. Consensus, PCR-clones, cell culture, DNA. [Murphy et al.(1993)]. GenBank accession numbers L11459–L11460, L11463–L11468, L11476, L11480–L11481, L11504–L11505, L11511–L11513, L11519–L11521 and U43137.
4) **CF2.ID#:** These three sequences were kindly provided prior to publication by Dr. M.P. Kieny of Transgene, Strasbourg Cedex, France. They are part of a set of 14 HIV-1 gp120 isolates from Bangui, Central African Republic. The gp120 proteins have been expressed in a hybrid gp160 vaccinia virus expression system. Year of isolation and health status of individuals from which the virus was isolated were not provided. Viruses were isolated in the CAR by C. Mathiot and B. You (Pasteur Inst., Bangui), grown by F. Barre-Sinoussi and A. Deslandres (Pasteur Inst., Paris), and cloned and sequenced by D. Schmitt and M.P. Kieny. Genbank accession numbers U43110, U43170 and U43173.
5) **CM.CA10:** A single E subtype sequence from a set of 17 sequences from a very diverse set of isolates from Yaounde and Douala, Cameroon. The sequences were derived from asymptomatic and symptomatic HIV infected individuals, specifically, CA10 was symptomatic. Virus was isolated by culture with donor PBMCs, and nested PCR amplified. A single clone was sequenced representing each HIV-1 isolate. [Nkengasong et al.(1994)]. GenBank accession numbers for the entire set of 17 sequences in this publication: X80438–X80454.
6) **GB1.11643:** This sequence is from the British isolate 94–11643. The sequence was determined from PCR-amplified lymphocyte DNA. The gag gene from this isolate was subtype A, as is the gag gene from all subtype E virus studied to date. The patient is thought to have contracted the virus in Thailand, but currently lives in the United Kingdom. [Arnold et al.(1995c)]. GenBank accession number U21109.
7) **JP1.ID#:** These three sequences are from one Japanese and two Thai individuals living in Japan, obtained by direct sequencing of PCR-amplified proviral DNA from peripheral blood mononuclear cells. [Weniger et al.(1994)]. NIH3J is from a male Japanese national. NIH2T and NIH4T are from Thai female prostitutes, living in Japan. GenBank accession numbers L32085–L32087.
8) **MY.92-14103:** This sequence is from a Thailand infant who was adopted by Malaysians, and now lives in Malaysia. PCR products amplified from uncultured PBMCs were directly sequenced. [Brown et al.(1996)]. This report also included subtypes B and C in Malaysia. GenBank accession number U65551.
9) **MY.1786:** This sequence is from a dried blood spot collected in 1992 from a Female STD patient in Myanmar. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. [Cassol et al.(1996)]. GenBank accession number U53309.
10) **NL.TH94037:** This sequence is from a recent immigrant to The Netherlands from Thailand. The blood sample was collected in 1994. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR product was directly sequenced. [Lukashov et al.(1996)]. GenBank accession number L76911.
11) **TH.93TH253:** This sequence is from a 21-year-old man from Chiang Mai, Thailand and was previously named CMU010 or 302053. The patient had end-stage AIDS. The mode and year of infection are unknown. 93th253 was isolated in 1993 and expanded in human PBMCs, then expanded in H9 cells, and followed by lambda cloning and sequencing. The complete genome has been sequenced. [Gao et al.(1996b)]. GenBank accession number U51189.

12) **TH.CM240:** This sequences is from a 21 year old asymptomatic man from northern Thailand. The route of infection is believed to be heterosexual transmission. The blood sample was collected in 1990. The patient's PBMSCs were codultured with stimulated donore PBMCs and proviral DNA was harvested for PCR amplification and sequening of a cloned full-length genome PCR product. [Carr et al.(1996)]. The complete genome is found in GenBank with accession number U54771.
13) **TH.T8178:** This sequence comes from a study of the genetic heterogeneity and epidemiological distribution of HIV1 in Thailand. The host was a female prostitute and the sequence was obtained from PCR amplified PBMC DNA. [Ou et al.(1993)]. GenBank accession number L19239. See also B_TH.T8174.
14) **TH.N764:** This sequence is from a survey of IV drug using prisoners in Thailand. 12 of 13 sequences from Thai prisoners were of subtype B; N764, from patient (THP13) represents the only subtype E sequence identified in this set, from a prisoner infected in 1989. The sequences were obtained from PCR amplified PBMC DNA. [Kalish et al.(1994)]. GenBank accession number U15588.
15) **TH1.ID#:** These twelve sequences are from a set of 23 individuals from Thailand. PCR-direct, peripheral blood PBMC DNA. Referred to as Thai subtype A in [Ou et al.(1992b)] and [Ou et al.(1993)]. (Published erratum appears in *Lancet* **342**:250 (1993).) GenBank accession numbers L07443–L07445, L07447–L07448, L07457–L07459 and L07461–L07464.
16) **TH2.ID#:** Six of these eight sequences are from 16 isolates from HIV seropositive individuals from Thailand. Sequences were from PCR products derived from co-cultured PBMC DNA. The full length envelope gene sequences are available. [McCutchan et al.(1992)]. Please note: the "TN-ID#" locus names in the database correspond to the McCutchan et al.'s "CM-ID#" isolates. GenBank accession numbers L03698–L03701 and L03703–L03704. The other two (TH238, TN240) are also from Thailand, DNA from PBMC. [Mascola et al.(1994)]. GenBank accession numbers L14571, L14572.
17) **TH3.W2TH-ID#:** Fifteen sequences from asymptomatic individuals from Thailand in 1992. Consensus, PCR-clones, cell-culture DNA and RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. [De Wolf et al.(1994)]; [Osmanov et al.(1994)]; and [Gao et al.(1994a)]. GenBank accession numbers U08810–U08811, U08825–U08836 and U08742–U08761. Entry with accession number U09131 is also TH_W2TH022.
18) **TH4.D-ID#:** These three sequences are part of a set of sequences generated for the DAIDS variation program in the laboratory of Dr. Beatrice Hahn at the University of Alabama. They are clones from expanded culture stocks, and are excised from full gp160 sequences. The sequence ID numbers are abbreviated, for example D3TH966 can be read as DAIDS sequence (D), isolated in 1993 (3), Thai (TH), patient 301966 (966). [Gao et al.(1996a)]. GenBank accession numbers U08456–U08458.
19) **TH5.ID#:** These seventeen consensus and five individual sequences are from twenty two patients with AIDS involved in a study of genotypic and phenotypic characteristics of Thai HIV-1. Blood samples were collected between July and December 1993. All sequences were derived from PCR amplified PBMC DNA, after patient PBMCs were cocultured with virus-free donor PBMCs. CMU01, CMU03, CMU04, CMU05, CMU07, and CMU10 are NSI, the rest are SI, as determined by syncytium formation in the cocultured cells. CM = Chaing Mai University Hospital. KH = Kavila Army Hospital. All subjects were males and reported past contact with commercial female sex workers, but no history of drug injection, blood transfusion or homosexual contact. [Yu et al.(1995)]. GenBank accession numbers U25550–U25626. Longer sequences from samples KH003, KH005, KH008, CMU02, CMU08 and CMU010 were determined in 1996 [McCutchan et al.(1996)]. GenBank accession numbers U48264–U48269.
20) **TH6.ID#:** These three sequences are E subtype sequences from Thailand. Two individuals believed to be dually infected with subtypes B and E were analyzed. It is not clear from the paper or the GenBank entries which sequences came from individual 1 and which from 2. [Artenstein et al.(1995)]. Genbank accession numbers U21472, U21474, U21476. See also B_TH5.ID#.
21) **TH7.ID#:** These two sequences are from samples collected in 1993 in Thailand. Patients 1018 and 1110 were asymptomatic. [McCutchan et al.(1996)]. GenBank accession numbers U48273–U48274.
22) **TH8.ID#:** These two sequences were from dried blood spots collected in 1992 from a heterosexual (0289) and an IV drug user (0103). DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. [Cassol et al.(1996)]. GenBank accession numbers U53312 and U53313.
23) **TH9.ID#:** These 14 sequences are from 84 IV drug users in Bangkok, Thailand, who were undergoing methadone treatment at 14 treatment clinics. Blood samples were collected between January and April,

1994. Uncultured PBMC DNA from each patient was PCR amplified, and the PCR product was directly sequenced. Of the 84 patients sampled, 69 were Thai B, one (091) was typical subtype B, and 14 were subtype E. [Kalish et al.(1995)]. GenBank accession numbers U22542, U22548, U22553, U22557, U22561, U22567, U22575, U22604, U22609, U22611, U22612, U22617, U22624 and U22625. See also B_TH7.

24) **TH10.ID#:** These sequences are from Thailand. They are part of several complete env gene sequences obtained for the World Health Organization. HMA subtyping as well as sequence-determined subtyping was done on each one. Sequences were PCR amplified from cocultured PBMCs. Two clones from each isolate were sequenced. [Penny et al.(1996)]. GenBank accession numbers U39256 and U39260, 92TH002; U39255 and U39261, 92TH011.
25) **TW1.ID#:** These 3 sequences are from healthy HIV-1 carriers or AIDS patients from Taiwan [Chang et al.(1996)]. Other subtypes found in Taiwan in this study were B, C, F and G. Genbank accession numbers U73060, U73062 and U73070.
26) **US.POC30506:** This sequence is from a U.S. serviceman who aquired an HIV-1 infection while deployed in Thailand. He was asymptomatic when the sample for this sequence was collected in 1993. [McCutchan et al.(1996)]. GenBank accession number U48272.
27) **UY.ID#:** These four sequences are from Uruguayan servicemen who aquired HIV-1 infections while deployed as United Nations peacekeepers in Cambodia in 1993. All four were asymptomatic when samples were collected for these sequences in 1993. [McCutchan et al.(1996)]. GenBank accession numbers U48275–U48278.
28) **VN1.ID#:** These 3 sequences are from South Vietnam. [Menu et al.(1996)]. The sequences were from IV drug users in Ho Chi Minh city and Dong Nai, and a female prostitute in Can Tho. A fourth sequence, from Ho Chi Minh city, was found to be subtype B. Genbank accession numbers U29206-U29208.
29) **VN2.ID#:** These 4 sequences are from South Viet Nam. VN1 and VN2 were from healthy 17 and 25 year old female prostitutes from Can Tho and An Giang. VN3 and VN 4 were from male IV drug users. VN3 was 43, had pruritus and splenomegaly, and was from Nha Trang. VN4 was 31, healthy and was also from Nha Trang. [Nerurkar et al.(1996)]. Genbank accession numbers U45239, U45240, U48719 and U48720.

NOTE:

1) While the sequences in this subtype were distinct over this region of env from the other four env subtypes, in the gag gene it is not possible to make a distinction between this subtype and subtype A. What this means is that the isolates for which both gag and env are sequenced which cluster together as the "A" subtype in gag, are very distinctive in env and are broken down into two subtypes. env "A" and env "E". This holds true for the E subtypes sequences that originated in Thailand, as well as the E subtype isolate from the Central African Republic for which gag sequence was obtained. [McCutchan et al.(1992)]; [Louwagie et al.(1993)]; and [Murphy et al.(1993)]. Complete genomes of subtype A and subtype E viruses became available in late 1996 [Gao et al.(1996b)], genbank accession numbers U51188–U51190.
2) The relative lack of diversity in the Thai sequences in this subtype relative to the other subtypes is likely to be a consequence of the short time span of the HIV-1 subtype E epidemic in Thailand. [McCutchan et al.(1992)], and [Ou et al.(1992b)].

F Subtype

At this time there are viral sequences from 59 HIV-1 infected individuals associated with HIV-1 subtype F. The F subtype consensus sequence (F_CONSENSUS_96) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published.

1) **AR1.ID#:** These two sequences are from direct sequencing of PCR products from uncultured PBMCs, from 1993 samples from Buenos Aires, Argentina. Patient 21280 had AIDS and reported IV drug abuse. Patient 20016 was asymptomatic and HIV risk behavior was unknown. Two other samples taken from unrelated patients in 1993 were subtypes B or B/F recombinant. [Marquina et al.(1996)]. GenBank accession numbers U68522 and U68524.
2) **AR2.ID#:** These 3 sequences are from Rosario, Argentina. A total of 24 patients from different risk groups visiting a clinic in Rosario were included in this study. Of the 14 sequences determined, 11 were found to belong to subtype B and 3 were found to belong to subtype F. DNA was extracted from whole blood and PCR amplified. PCR products were directly sequenced. Subtypes of all 24 patients were also tested by HMA. [Campodonico et al.(1996)]. GenBank accession numbers U37032, U37033 and U37043.
3) **BR.7944:** This sequence represents a single env F subtype sequence found among 22 Brazilian outpatients with varying degrees of disease progression. It was identified by Potts et al. as the single sequence which did not cluster with North American sequences in phylogenetic analysis. Consensus, PCR clones, peripheral blood PBMC DNA. [Potts et al.(1993b)]. GenBank accession number L19237.
4) **BR.RJI03:** An F subtype sequence from Rio de Janeiro, Brazil. 26 additional B and a B-F recombinant were also observed in this set. Year of isolation for RJI03 was 1993, from a woman of CDC clinical stage II. [Morgado et al.(1994)]. DNA was amplified directly from PBMCs of an HIV infected woman with CDC stage II disease, and the PCR product was directly sequenced. GenBank accession number U00422. See also [Sabino et al.(1994c)] GenBank accession number U08974.
5) **BR1.BZ-ID#:** Three sequences from Brazil of the F subtype. Full length env (gp160) was amplified from proviral DNA of cultured PBMCs, cloned and sequenced. [Louwagie et al.(1994)]. GenBank accession numbers L22082, L22084 and L22085. The gag gene of these same isolates is found in L22083, L22086 and L11751. The gag region of BZ126 seems to be a subtype A outlier with a strong similarity to subtype C at the 3' end. See also B_BR4.BZ-ID#.
6) **BR2.HSP#:** These 5 sequences are from Sao Paulo Brazil. Although these sequences are listed as unpublished in the database, they seem to be an extension of work published in [Morgado et al.(1994)] and [Sabino et al.(1994c)]. GenBank accession numbers U31588, U31592–U31595.
7) **BR3.93BR029:** This sequence is from Brazil. It is one of several complete env gene sequences obtained for the World Health Organization. HMA subtyping as well as sequence-determined subtyping was done on each one. Sequences were PCR amplified from cocultured PBMCs. Two to three clones from each isolate were sequenced. [Penny et al.(1996)]. Another env sequence from this same patient was determined by a second group. [Gao et al.(1996a)]. GenBank accession numbers U27413, U39235 and U39236.
8) **CM.CA-ID#:** These sequences are 3 of 17 sequences from a very diverse set of isolates from Yaounde and Douala, Cameroon. The sequences were derived from asymptomatic and symptomatic HIV infected individuals; specifically patients CA16 and CA20 were asymptomatic and patient CA4 was symptomatic. Virus was isolated by culture with donor PBMCs, and nested PCR amplified. A single clone was sequenced representing each HIV-1 isolate. The F subtype designation of these sequences is tentative. Although the F subtype sequences from Cameroon and Brazil consistently form a clade in phylogenetic analyses, the branch lengths between isolates from the two countries are typical of inter-subtype distances, and sequences from the two countries each form their own distinct clade within the F subtype (HIV database and Wouter Janssens, personal communication). [Nkengasong et al.(1994)]. GenBank accession numbers for the entire set of 17 sequences studied in this publication: X80438–X80454.
9) **CY.HO44-1:** This is a single sequence from two individuals who were heterosexual partners of one another. Patient 16 was a 29 year old bisexual male who was born and lived in Zaire, before moving to Cyprus. He was symptomatic with a CD4 count of 60 and had been seropositive for at least 6 years. Patient 44 was a 32 year old heterosexual female. She was asymptomatic with a CD4 count of 1,136. These samples, like others in this study (see also subtypes A, B, C, and I) were collected in February 1994

from the AIDS clinic in Nicosia, Cyprus. DNA was extracted from patient PBMCs and PCR amplified. After a second round of PCR, products were cloned and sequenced. One clone from patient 16 and one from patient 44 were sequenced. [Kostrikis et al.(1995)]. Because of the close epidemiological linkage, only the clone from patient 44 is presented here. Genbank accession numbers U28662 (16) and U28679 (44).

10) **GA1.VI354:** This sequence is from a 1989 sample from a patient with AIDS living in Libreville, Gabon. Method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced. [Delaporte et al.(1996)]. GenBank Accession number X90923. See also subtypes A, C, D, G and O sequences from this same study.
11) **NL1.ID#:** These 3 sequences are from recent immigrants to The Netherlands from Brazil and Zaire. The fist two letters of the ID# represent the two letter country code for the previous residence of the patient. The next two numerals represent the year of isolation. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced. [Lukashov et al.(1996)]. GenBank accession numbers L76899, ZR9306911; L76871, ZR890819; L76901, BR9400960.
12) **RO1.ID#:** These nine sequences are from isolates from Romanian children, in different clinical stages. All isolates showed cytopathic properties in peripheral blood mononuclear cells. DNA sequences are direct sequences of PCR products amplified from co-cultured PBMCs. The patients are also known as RM(A-J). [Dumitrescu et al.(1994)]. GenBank accession numbers L19571–L19579.
13) **RO2.RM-ID#:** These 24 sequences are from isolates from orphaned Romanian children, ranging in age from 2.5 to 6 years, admitted to a clinic in Tirgu Mures, Romania. All children were referred to this clinic with serious infections and are believed to have been infected horizontally in different orphanages. Virus was isolated after coculture with donor PBMCs. Proviral DNA from cocultured PBMCs was PCR amplified and the PCR products were directly sequenced. [Holm-Hansen et al.(1995)]. GenBank accession numbers X77964–X77987.
14) **TW1.252:** This sequence is one of a set of sequences from Taiwan. Other sequences in the set were subtypes B, C, E or G. [Chang et al.(1996)]. GenBank accession number U67765.

G Subtype

At this time there are viral sequences from 23 HIV-1 infected individuals associated with HIV-1 subtype G. The G subtype consensus sequence (G_CONSENSUS_96) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published and/or have been made available for printing in the database by their authors.

1) **BJ1.ID#:** These 2 sequences are from female prostitutes, born in either Ghana or Togo, who live in Benin. 43 is from directly sequenced PCR product, derived via RT-PCR from patient serum RNA. 259 is from cloned PCR product, also by RT-PCR from serum RNA. [Heyndrickx et al.(1996)]. GenBank accession numbers U61872 and U61874. Subtype A sequences were also determined in this study.
2) **CF.4067:** This sequence was associated with the C subtype in first analysis of the C2V3 region ([Murphy et al.(1993)]), but when a full gp120 sequence became avaliable from this isolate, and phylogenetic analysis was performed including some of the new subtype G sequences, it was more closely associated with G. The full length sequence was kindly provided prior to publication by Dr. MP Kieny of Transgene, Strasbourg Cedex, France. It is part of a set of 14 HIV-1 gp120 isolates from Bangui, Central African Republic. The gp120 proteins have been expressed in a hybrid gp160 vaccinia virus expression system. Year of isolation and health status of individuals from which the virus was isolated were not provided. Viruses were isolated in the CAR by C Mathiot and B You (Pasteur Inst., Bangui), grown by F Barre-Sinoussi and A Deslandres (Pasteur Inst., Paris), and cloned and sequenced by D Schmitt and MP Kieny. GenBank accession numbers L11499 and L11500, U43169.
3) **GA.LBV21-7:** A sequence from Gabon from a set of HIV-1 viral isolates from Africa. This sequence was derived from a clone of PCR amplified DNA from cultured PBMCs. It represents a fragment of a full length env sequence. [Janssens et al.(1994b)]. GenBank accession number U09664.
4) **GA.VI525:** A sequence from Gabon from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1994)] and [Janssens et al.(1994b)]. GenBank accession numbers L22953 and U09665. The same isolate was classified as subtype H in gag, [Louwagie et al.(1993)], GenBank accession number L11792.
5) **GA1.ID#:** These 2 sequences are from Gabon. G98 is from a 1988 or 1989 sample from a patient with AIDS living in Franceville, Gabon who moved there from Niger. VI526 is from a 1990 sample from an AIDS patient at the Libreville General Hospital. Method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced. [Delaporte et al.(1996)]. GenBank Accession numbers X90916, G98; X90922, VI526. See also subtypes A, C, D, F and O sequences from this same study.
6) **GB1.22:** This sequence is a consensus of three clones from an infected infant in a mother-infant transmission study. The sequences were obtained via PCR from cell lysates, with sequencing of cloned PCR products. The infant was 3 months old at the time of blood drawing, and had pneumonia. [Arnold et al.(1995b)]. GenBank accession numbers U26304–U26306. Envelope sequences for the mother are found in GenBank entries U26301–U26302, and gag sequences for mother and infant are in U26303 and U26307.
7) **KP.Kr121** This sequence is from an unpublished GenBank entry with accession number X93469.
8) **NG1.ID#:** These four sequences are G subtype sequences from Nigeria [Abimiku et al.(1994)]. JP882 and JV832 were derived from AIDS patients, and G3 and G9 from healthy women. G9 was cultured on the T cell line CEM-SS, and the other three isolates were cocultured with uninfected donor PBMCs. DNA from viral cultures was PCR amplified, cloned and sequenced. GenBank accession numbers U13208–U13209, U13211 and U13213.
9) **NL.127C** This consensus sequence represents sequences generated from PCR amplified plasma RNA from one of three infants in a Dutch mother/infant study. A sample was collected from the infant at 1.5 months of age. Samples were also collected from the mother before birth, at birth and after birth. Mother sequences are not included in this consensus. [Mulder-Kampinga et al.(1993)]. [Mulder-Kampinga et al.(1995)]. Infant 127 sequences are from GenBank accession numbers Z47817–Z47832. Mother 127 sequences are from GenBank accession numbers Z47833–Z47880. Gag gene sequences from mother/child pairs are also available in Genbank accession numbers Z47903–Z47911; Z47912–Z47928;

Z47929–Z47935; Z47936–Z47950. The second mother/child pair was also from the Netherlands, see B_NL.114C. The third mother/infant pair in this study was from Rwanda, see A_RW.564C.

10) **NL1.ID#:** These 4 sequences are from recent immigrants to The Netherlands from Brazil and Zaire. The fist two letters of the ID# represent the two letter country code for the previous residence of the patient. The next two numerals represent the year of isolation. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced. [Lukashov et al.(1996)]. GenBank accession numbers L76880, ZR911976; L76906, LR9401885; L76884, UM9210113; L76902, GH9401230.
11) **RU.RUS12A:** This sequence is from [Lukashov et al.(1995)]. GenBank accession number L38413.
12) **RU1.ID#:** This sequence is from Russia. Bobkov et al. 1996 Unpublished. GenBank accession numbers U33095, U33096.
13) **TW1.267:** This sequence is from directly sequenced PCR product from uncultured PBMCs from Taiwan. [Chang et al.(1996)]. Most of the Taiwanese sequences determined in this study were subtype B, but subtypes C, E and F were also found. GenBank accession number U73058.
14) **UG.JW3:** This single sequence of G HIV1 is from a female patient with stage IV disease and CD4 count of 20, from Uganda. The patient had recently migrated to the United Kingdom from Uganda, but contracted the HIV in Uganda. The sequence is referred to as JW3 in [Kaleebu et al.(1995)] but was previously referred to as K1, by the HIV database. It has been given the name 92UG975.10 by the World Health Organization. GenBank accession numbers U22010, U27426.

H Subtype

At this time there are viral sequences from 2 HIV-1 infected individuals associated with HIV-1 subtype H. The H subtype consensus sequence (H_CONSENSUS_96) generated from these 2 sequences was based on the most common amino acid found in each position of an alignment. Both of these sequences have been published and/or have been made available for printing in the database by their authors. Eight sequences that are too short for classification are closer to H than to other subtypes. The locus names (ID's) and sources of the sequences are:

1) **CM.CA13:** A sequence from Cameroon from a set of HIV-1 viral isolates from Africa used to define the prototype G and H env sequences. This sequence was derived from a clone of PCR amplified DNA from cultured PBMCs. It represents a fragment of a 900 base pair sequence. [Janssens et al.(1994d)] and [Nkengasong et al.(1994)]. The H subtype association is not always clearly apparent using some sets of background sequences for comparison, and neighbor joining trees (HIV database, Wouter Janssens, personal communication), although parsimony trees confirmed the original association documented in [Janssens et al.(1994b)]. GenBank accession number U09667.
2) **ZR.VI557:** A sequence from Zaire from a set of HIV-1 viral isolates from Africa used to define the prototype G and H env sequences. This sequence was derived from a clone of PCR amplified DNA from cultured PBMCs. It represents a fragment of a 900 base pair sequence. [Janssens et al.(1994b)]. GenBank accession number U09666.

I Subtype

At this time there are viral sequences from 2 HIV-1 infected individuals associated with HIV-1 subtype I. The subtype I consensus sequence reflects the sequencing of two clones from one individual and one clone from the other. Both of these sequences have been published.

1) **CY.HOcon:** This is a consensus sequence from two individuals who were heterosexual partners of one another, and former IV drug users. They had lived for several years in Athens, Greece as well as in Cyprus. These samples, like others in this study (see also subtypes A, B, C, and F) were collected in February 1994 from the AIDS clinic in Nicosia, Cyprus. Patient HO31 was a 24 year old asymptomatic female known to have been HIV seropositive for at least 5 years. Patient HO32 was a 35 year old asymptomatic male, also seropositive for at least 5 years. DNA was extracted from patient PBMCs and PCR amplified. After a second round of PCR, products were cloned and sequenced. Two clones from HO32 and one from HO31 were sequenced. [Kostrikis et al.(1995)]. Genbank accession numbers U28672, U28673 and U28685.

J Subtype

At this time there are viral sequences from 5 HIV-1 infected individuals associated with HIV-1 subtype J. The J subtype consensus sequence (J_CONSENSUS_96) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published or made available to the database for printing.

1) **SE1.SE#:** Two sequences from Sweden, both from patients who were recent immigrants from Zaire. [Leitner et al.(1995)]. Sample 2022 as collected in December 1993 from an asymptomatic female with a CD4 count of 184. Her first know seropositive sample is from May 1990, but epidemiological investigation indicated that she was infected in Zaire between 1981 and 1986. Sample 7887 was collected in October 1994 from an asymptomatic male who had tested seronegative in Sweden in January 1993, and who had a seropositive sample in August 1994. His CD4 count was normal at 567. Both patients were heterosexual and a thorough epidemiological investigation revealed no contact or shared contacts between the two. The two sequences were 95% identical to each other over 255 bases of env and 98% identical to each other over 460 bases of gag. GenBank accession numbers L41176 and L41177. Gag gene sequences from these same individuals are in L41178 and L41179.
2) **GM1.GM#:** Three sequences from Gambia, as yet unpublished, which seem to cluster with the two Swedish sequences by Leitner. GenBank accession numbers U33099, U33100, U33102. Bobkov et al. 1996 unpublished. See also Gambian sequences of subtypes B and C.

O Subtype

At this time there are viral sequences 13 HIV-1 infected individuals associated with HIV-1 subtype O that have been published and/or have been made available for printing in the database by their authors. The O subtype consensus sequence (O_CONSENSUS_96) generated from these sequences was based on the most common amino acid found in each position of the alignment; when there was no consensus in a position an "X" was used. These sequences represent a set of sequences that are extremely divergent relative to other HIV-1's. The subtypes A-H have been grouped together under the heading "M" for main. "O" sequences are as different from one another as are sequences from different "M" subtypes. A more complete discussion of O-group viral sequences is also included in this section of the Compendium (Korber, B.T., Loussert-Ajakai, I. Blouin, J. and Saragosti, S. 1996 A Comparison of HIV-1 Group M and Group O Functional and Immunogenic Domains in the Gag p24 Protein and the C2V3 Region of the Envelope Protein).

1) **CM.CA9:** This sequence is from an individual living in Cameroon. [Janssens et al.(1994e)]. No GenBank entry is yet available. The pol gene from this isolate has GenBank accession number X78476.
2) **FR.DUR:** This sequence is as yet unpublished, by J.H.M. Cohen et al. GenBank accession number X84327.
3) **GA1.VI686:** This sequence is from a 1992 sample from a Gabonese woman with AIDS, taken at the Libreville General Hospital in Gabon. Method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced. [Janssens et al.(1994e)], [Delaporte et al.(1996)]. GenBank accession number X96526. The pol gene from this isolate has GenBank accession number X78477. See also subtypes A, C, D, F and G sequences from this same study.
4) **CM.ANT70:** The complete viral genome has been sequenced from this viral isolate derived from a symptomatic Cameroonian, CDC stage III. [R. et al.(1990)] and [Vanden Haesevelde et al.(1994)]. GenBank accession number M31171. LTR and partial env sequences were also presented in L20587 and L23119.
5) **CM.MVP5180:** The complete viral genome has been sequenced from an isolate derived from a Cameroonian woman, sampled in 1991; the donor died of AIDS in 1992. The viral isolate MVP-5180 was grown in several human T-cell lines and the monocytic U937 line. [Gurtler et al.(1994)]. GenBank accession number L20571.
6) **ES.1158:** This sequence was from a 35 year old man from Spain. Two blood samples from this same man were collected in April and September 1995. The V3 region was PCR amplified from uncultured PBMCs, cloned into pGEM-5ZF and an indiviudal clone sequenced. The April sequence is shown here. The sequence from the September blood sample (681, GenBank accession number U62617) is also available. GenBank accession number U62618.
7) **FR.VAU:** This sequence was derived from an isolate from a French woman who died of AIDS in 1992. DNA was extracted from VAU infected PBMCs, PCR amplified, cloned, and gp160 env was sequenced. The viral isolate was highly cytopathic. [Charneau et al.(1994)]. GenBank accession number X80020.
8) **FR.CF#:** These seven consensus sequences are from Cameroonian patients living in France. [Loussert-Ajaka et al.(1995)]. PBMC proviral DNA was PCR amplified and 3-6 clones from each patient were sequenced. The consensus of the 3-6 clones is presented. GenBank accession numbers U24562–U24568. Gag gene sequences for these patients are also available with GenBank accession numbers U24706–U24712.

Uncertain Classification

At this time there are viral sequences from 15 HIV-1 infected individuals that are not clearly associated with any of the HIV-1 genetic subtypes A through J. They either appeared distinct from the subtypes A-J in phylogenetic analysis, or else the subtype association was unclear, with different associations in different analyses. For some of the shorter gene fragments, subtype associations might have been established if more sequence information was available or if a different set of sequences was included in the background set used to define subtype associations. Some of these sequences may be representatives of subtypes as divergent as A-J, but only a single limited sample is yet available. Still others may represent recombinant genomes.

1) **AR.20021:** This B/F recombinant sequence is from direct sequencing of PCR product from uncultured PBMCs, from a 1993 sample from Buenos Aires, Argentina. The patient was asymptomatic and HIV risk behavior was unknown. Three other samples taken from unrelated patients in 1993 were subtypes F (2) or B (1). [Marquina et al.(1996)]. GenBank accession number U68523.
2) **BR3.RJI01** This sequence is B-F recombinant in the V3 region. DNA was amplified directly from PBMCs of an HIV infected woman with CDC stage II/A disease in August, 1992, and the PCR product was directly sequenced [Morgado et al.(1994)]. More V3 region sequences from this individual (RJ549 from April 1992) and her sexual partner (RJ548 from April 1992) were also sequenced [Sabino et al.(1994c)]. GenBank accession numbers U00420, U08953–U08955, U08957–U08960, U08962–U08964, U10019–U10029, U08972, U08973 and U08965–U08971.
3) **BU1.91BU009:** This C/D sequence is from Burundi. It is one of several complete env gene sequences obtained for the World Health Organization. HMA subtyping as well as sequence-determined subtyping was done on each one. Sequences were PCR amplified from cocultured PBMCs. Two to three clones from each isolate were sequenced. 91BU009 groups with subtype D in a neighbor- joining tree of the V3 region and with subtype C in other regions. [Penny et al.(1996)]. GenBank accession numbers U39253 and U39254.
4) **CF1.ID#:** These four sequences are from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990–1991. Consensus, PCR-clones, cell culture, DNA. [Murphy et al.(1993)]. GenBank accession numbers L11482–L11483, L11497, L11508–L11510 and L11514–L11515. Janssens et al. classified 4056 as an H subtype sequence [Janssens et al.(1994d)]. D. Schmitt provided an unpublished sequence of CF.4081, U43174.
5) **KE.K124:** A Kenyan sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. This isolate is not clearly associated with D subtype; however, Louwagie and colleagues found that it associated with the D subtype in env, and the A subtype in gag. Using parsimony analysis, we found that it was difficult to determine a clear association, and this observation was confirmed by Wouter Janssens (personal communication). GenBank accession number L22942.
6) **KE.KEN976** This is a single unclassified sequence from a set of patients who were part of a 1990–1992 cohort study of maternal risk factors in mother to child transmission, including 22 pregnant women and an infant from Kenya. The C2V3 region was sequenced. [Janssens et al.(1994a)]. GenBank accession number U12992.
7) **NL.RW94028:** This sequence is from a recent immigrant to The Netherlands from Rwanda. The blood sample was collected in 1994. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR product was directly sequenced. [Lukashov et al.(1996)]. GenBank accession number L76909.
8) **UG1.ID#:** Two clones from each of these two D/A recombinant Ugandan isolates were sequenced, but only the sequences of 92UG035 clone 21 and C6080 clone 09 are shown here. [Douglas et al.(1996)]. The publication shows C6080 as subtype A, but analyses done at Los Alamos indicate that it is a D/A recombinant, like 92UG035. Other Ugandan isoaltes sequenced in this study were subtype D. London subtype B clones were also reported. Complete envelope gp160 sequences were reported for all isolates. GenBank accession numbers U36865, U36866, U36881 and U36883.
9) **SE.KI4803:** This sequence is from patient number 24 described in [Asjo et al.(1986)]. Several molecular clones from this patient have been extensively characterized in [Tan et al.(1993)]. Complete env gp120

sequences for 8 clones were determined in [McKeating et al.(1996)] and one of the 8 sequences (clone 13) is presented here. Genbank accession numbers for 7 if the 8 clones (clone 32 does not apear in GenBank) are U57788–U57794.

10) **ZM.ZAM184:** This Zambian sequence is an outlier, though in some phylogenetic analysis it appears most closely associated with the A subtype. In particular it is closely associated with A_CF.SAS U43171 (100/100 replicates in parsimony analysis of gp120). Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession number L22955.
11) **ZR.Z3:** This sequence is from the 1983 Zairean isolate Z-3 (non-infectious, possibly due to frame-shift). [Willey et al.(1986)]. GenBank accession number K03347.

References

[Abele & DeBry(1992)] L. G. Abele & R. W. DeBry. Florida dentist case: research affiliation and ethics. *Science* **255**:903, 1992. OTE: (Medline: 92188144).

[Abimiku et al.(1994)] A. G. Abimiku, T. L. Stern, A. Zwandor, P. D. Markham, C. Calef, S. Kyari, W. C. Saxinger, R. C. Gallo, M. Robert-Guroff, & M. S. Reitz. Subgroup G HIV type 1 isolates from Nigeria. *AIDS Res Hum Retroviruses* **10**:1581–1583, 1994. OTE: (Medline: 95194721).

[Adachi et al.(1986)] A. Adachi, H. E. Gendelman, S. Koenig, T. Folks, R. Willey, A. Rabson, & M. A. Martin. Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. *J Virol* **59**:284–91, 1986. OTE: (Medline: 86281827), (Genbank: M19921).

[Ahmad et al.(1995)] N. Ahmad, B. M. Baroudy, R. C. Baker, & C. Chappey. Genetic analysis of human immunodeficiency virus type 1 envelope V3 region isolates from mothers and infants after perinatal transmission. *J Virol* **69**:1001–1012, 1995. OTE: (Medline: 95115054), (Genbank: U16390 U16391 U16392 U16393 U16394 U16395 U16396 U16397 U16398 U16399 U16400 U16401 U16402 U16403 U16404 U16405 U16406 U16407 U16408 U16409 U16410 U16411 U16412 U16413 U16414 U16415 U16416 U16417 U16418).

[Ait-Khaled & Emery(1993)] M. Ait-Khaled & V. C. Emery. Sequence variation within the human immunodeficiency virus V3 loop at seroconversion. *J Med Virol* **41**:270–274, 1993. OTE: (Medline: 94149438).

[Albert et al.(1992)] J. Albert, L. Franzen, M. Jansson, G. Scarlatti, P. K. Kataaha, E. Katabira, F. Mubiro, M. Rydaker, P. Rossi, & U. P. et al. Ugandan HIV-1 V3 loop sequences closely related to the U.S./European consensus. *Virology* **190**:674–81, 1992. OTE: (Medline: 92391084), (Genbank: M98894 M98895 M98896 M98897 M98898 M98899 M98900 M98901 M98902 M98903 M98904 M98905 M98906 M98907 M98908 M98909 M98910 M98911 M98912 M98913 M98914 M98915 M98916 M98917 M98918 M98919 M98920 M98921 M98922 M98923).

[Alizon et al.(1986)] M. Alizon, S. Wain-Hobson, L. Montagnier, & P. Sonigo. Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. *Cell* **46**:63–74, 1986. OTE: (Medline: 86245056), (Genbank: K02013 K03454 K03456).

[Anand et al.(1989)] R. Anand, R. Thayer, A. Srinivasan, S. Nayyar, M. Gardner, P. Luciw, & S. Dandekar. Biological and molecular characterization of human immunodeficiency virus (HIV-1BR) from the brain of a patient with progressive dementia. *Virology* **168**:79–89, 1989. OTE: (Medline: 89085613), (Genbank: M21098).

[Andeweg et al.(1992)] A. C. Andeweg, M. Groenink, P. Leeflang, G. de R. E., A. D. Osterhaus, M. Tersmette, & M. L. Bosch. Genetic and functional analysis of a set of HIV-1 envelope genes

obtained from biological clones with varying syncytium-inducing capacities. *AIDS Res Hum Retroviruses* **8**:1803–13, 1992. OTE: (Medline: 93090463).

[Antonioli et al.(1995)] I. M. Antonioli, C. Baumberger, S. Yerly, & L. Perrin. V3 sequences in primary HIV-1 infection. *AIDS* **9**:11–17, 1995. OTE: (Medline: 95200567).

[Arnold et al.(1995a)] C. Arnold, P. Balfe, & J. P. Clewley. Sequence distances between env genes of HIV-1 from individuals infected from the same source: implications for the investigation of possible transmission events. *Virology* **211**:198–203, 1995a. OTE: (Medline: 95373135), (Genbank: U23112 U23113 U23114 U23115 U23116 U23117 U23118 U23119 U23120 U23121 U23122 U23123 U23124 U23125 U23126 U23127 U23128 U23129 U23130 U23131 U23132 U23133 U23134 U23135 U23136 U23137 U23138).

[Arnold et al.(1995b)] C. Arnold, K. L. Barlow, S. Kaye, C. Loveday, P. Balfe, & J. P. Clewley. HIV type 1 sequence subtype G transmission from mother to infant: failure of variant sequence species to amplify in the Roche Amplicor Test. *AIDS Res Hum Retroviruses* **11**:999–1001, 1995b. OTE: (Medline: 96020103), (Genbank: U26301 U26302 U26303 U26304 U26305 U26306 U26307).

[Arnold et al.(1995c)] C. Arnold, K. L. Barlow, J. V. Parry, & J. P. Clewley. At least five HIV-1 sequence subtypes (A, B, C, D, A/E) occur in England. *AIDS Res Hum Retroviruses* **11**:427–429, 1995c. OTE: (Medline: 95306148), (Genbank: U21095 U21096 U21097 U21098 U21099 U21100 U21106 Z33479).

[Artenstein et al.(1995)] A. W. Artenstein, T. C. VanCott, J. R. Mascola, J. K. Carr, P. A. Hegerich, J. Gaywee, E. Sanders-Buell, M. L. Robb, D. E. Dayhoff, & S. T. et al. Dual infection with human immunodeficiency virus type 1 of distinct envelope subtypes in humans. *J Infect Dis* **171**:805–810, 1995. OTE: (Medline: 95221984), (Genbank: U21471 U21472 U21473 U21474 U21475 U21476).

[Ashkenazi et al.(1991)] A. Ashkenazi, D. H. Smith, S. A. Marsters, L. Riddle, T. J. Gregory, D. D. Ho, & D. J. Capon. Resistance of primary isolates of human immunodeficiency virus type 1 to soluble CD4 is independent of CD4-rgp120 binding affinity. *Proc Natl Acad Sci U S A* **88**:7056–60, 1991. OTE: (Medline: 91334404).

[Asjo et al.(1986)] B. Asjo, L. Morfeldt-Manson, J. Albert, G. Biberfeld, A. Karlsson, K. Lidman, & E. M. Fenyo. Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection. *Lancet* **2**:660–662, 1986. OTE: (Medline: 87013480).

[Atkin et al.(1993)] A. Atkin, G. Pestano, D. Serwadda, A. M. Prince, D. Pascual, N. Sewankambo, & W. M. Boto. Phylogenetic and serological characterization of two Ugandan HIV-1 isolates. *AIDS Res Hum Retroviruses* **9**:351–6, 1993. OTE: (Medline: 93290930), (Genbank: M98503 M98504).

[Audoly et al.(1996)] G. Audoly, S. Fillon, P. Pinay, & C. Desgranges. In vivo and in vitro sequence diversity of the V3 env region of HIV type 1 from the Ivory Coast. *AIDS Res Hum Retroviruses* **12**:1073–1075, 1996. OTE: (Medline: 96424763).

[Ayyavoo et al.(1996)] V. Ayyavoo, K. E. Ugen, L. S. Fernandes, J. J. Goedert, A. Rubinstein, W. V. Williams, & D. B. Weiner. Analysis of genetic heterogeneity, antigenicity, and biological characteristics of HIV-1 in a maternal transmitter and nontransmitter patient pair. *DNA Cell Biol.* **15**:571–580, 1996. OTE: (Medline: 96326292), (Genbank: U07833 U07834 U07835 U07836 U07837 U07839 U07840 U07841 U07891 U07892 U07893 U07894 U07895 U07896 U07897 U07898 U07899 U07900 U07901 U07902 U07903 U07904 U07905 U07906 U07907 U07908 U07909 U07910 U07911 U07912).

[Balfe et al.(1990)] P. Balfe, P. Simmonds, C. A. Ludlam, J. O. Bishop, & A. J. Brown. Concurrent evolution of human immunodeficiency virus type 1 in patients infected from the same source: rate of sequence change and low frequency of inactivating mutations. *J Virol* **64**:6221–33, 1990. OTE: (Medline: 91056593), (Genbank: M36997).

[Baskar et al.(1994)] P. V. Baskar, S. C. Ray, R. Rao, T. C. Quinn, J. E. Hildreth, & R. C. Bollinger. Presence in India of HIV type 1 similar to North American strains. *AIDS Res Hum Retroviruses* **10**:1039–1041, 1994. OTE: (Medline: 95110628), (Genbank: L29091 L29092 L29093 L29094).

[Becker et al.(1995)] M. L. Becker, J. D. G., & W. B. Becker. Analysis of partial gag and env gene sequences of HIV type 1 strains from southern Africa. *AIDS Res. Hum. Retroviruses* **11**:1265–1267, 1995. OTE: (Medline: 96157217), (Genbank: X52154).

[Bex et al.(1994)] F. Bex, P. Hermans, S. Sprecher, A. Achour, R. Badjou, C. Desgranges, J. Cogniaux, P. Franchioli, C. Vanhulle, & A. Lachgar. Syngeneic adoptive transfer of anti-human immunodeficiency virus (HIV- 1)-primed lymphocytes from a vaccinated HIV-seronegative individual to his HIV-1-infected identical twin. *Blood* **84**:3317–26, 1994. OTE: (Medline: 95036224).

[Briant et al.(1995)] L. Briant, C. M. Wade, J. Puel, A. J. Brown, & M. Guyader. Analysis of envelope sequence variants suggests multiple mechanisms of mother-to-child transmission of human immunodeficiency virus type 1. *J Virol* **69**:3778–3788, 1995. OTE: (Medline: 95264471), (Genbank: U24717 U24718 U24719 U24720 U24721 U24722 U24723 U24724 U24725 U24726 U24727 U24728 U24729 U24730 U24731 U24732 U24733 U24734 U24735 U24736 U24737 U24738 U24739 U24740 U24741 U24742 U24743 U24744 U24745 U24746).

[Brown et al.(1996)] T. M. Brown, K. E. Robbins, M. Sinniah, T. S. Saraswathy, V. Lee, L. S. Hooi, B. V. ans C. C. Luo, C. Y. Ou, J. Rapier, G. Schochetman, & M. L. Kalish. Human immunodeficiency virus type 1 subtypes in Malaysia include B C, and E. *AIDS Res. Hum. Retroviruses* **12**:1655–1657, 1996.

[Bruce et al.(1993)] C. Bruce, C. Clegg, A. Featherstone, J. Smith, & J. Oram. Sequence analysis of the gp120 region of the env gene of Ugandan human immunodeficiency proviruses from a single individual. *AIDS Res Hum Retroviruses* **9**:357–63, 1993. OTE: (Medline: 93290931).

[Buonaguro et al.(1995)] L. Buonaguro, G. D. E., M. Monaco, D. Greco, P. Corti, E. Beth-Giraldo, F. M. Buonaguro, & G. Giraldo. Heteroduplex mobility assay and phylogenetic analysis of V3 region sequences of human immunodeficiency virus type 1 isolates from Gulu, northern Uganda. The Italian-Ugandan Cooperation AIDS Program. *J Virol* **69**:7971–7981, 1995. OTE: (Medline: 96079047).

[Burger et al.(1991)] H. Burger, B. Weiser, K. Flaherty, J. Gulla, P. N. Nguyen, & R. A. Gibbs. Evolution of human immunodeficiency virus type 1 nucleotide sequence diversity among close contacts. *Proc Natl Acad Sci U S A* **88**:11236–40, 1991. OTE: (Medline: 92107924), (Genbank: M77229 M77230 M77278).

[Cabello et al.(1995)] A. Cabello, M. Cabral, M. E. Vera, R. Kiefer, R. M. Azorero, J. Eberle, L. Gurtler, & B. V. A. Analysis of the V3 loop sequences from 10 HIV type 1-infected AIDS patients from Paraguay. *AIDS Res. Hum. Retroviruses* **11**:1135–7, 1995. OTE: (Medline: 96089222).

[Calabro et al.(1995)] M. L. Calabro, C. Zanotto, F. Calderazzo, C. Crivellaro, M. D. A., R. D. A., & L. Chieco-Bianchi. HIV-1 infection of the thymus: evidence for a cytopathic and thymotropic viral variant in vivo. *AIDS Res Hum Retroviruses* **11**:11–19, 1995. OTE: (Medline: 95251923).

[Campodonico et al.(1996)] M. Campodonico, W. Janssens, L. Heyndrickx, K. Fransen, A. Leonaers, F. F. Fay, M. Taborda, G. Van der Groen, & O. H. Fay. HIV type 1 subtypes in Argentina and genetic heterogeneity of the V3 region. *AIDS Res. Hum. Retroviruses* **12**:79–81, 1996. OTE: (Medline: 96423022).

[Carr et al.(1996)] J. K. Carr, M. O. Salminen, C. Koch, D. Gotte, A. W. Artenstein, P. A. Hegerich, L. S. D., D. S. Burke, & F. E. McCutchan. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J. Virol.* **70**:5935–5943, 1996. OTE: (Medline: 96323109), (Genbank: U54771).

[Cassol et al.(1996)] S. Cassol, B. Weniger, P. Babu, M. Htoon, A. Delaney, M. O'Shaughnessy, & C.-Y. Ou. Detection of hiv-1 env subtypes a, b, c, and e in asia using dried blood spots: A new surveillance tool for molecular epidemiology. *AIDS Res. Hum. Retroviruses* **12**:1435–1441, 1996.

[Chang et al.(1996)] K. Chang, C. Lin, J. Chen, C. Shih, H. C. Lin, R. Lin, S. Twu, & M. Salminen. Hiv-1 env gene diversity detected in taiwan. *AIDS Res. Hum. Retroviruses* **In Press**, 1996.

[Charneau et al.(1994)] P. Charneau, A. M. Borman, C. Quillent, D. Guetard, S. Chamaret, J. Cohen, G. Remy, L. Montagnier, & F. Clavel. Isolation and envelope sequence of a highly divergent HIV-1 isolate: definition of a new HIV-1 group. *Virology* **205**:247–253, 1994. OTE: (Medline: 95065659).

[Cheng-Mayer et al.(1988)] C. Cheng-Mayer, J. Homsy, L. A. Evans, & J. A. Levy. Identification of human immunodeficiency virus subtypes with distinct patterns of sensitivity to serum neutralization. *Proc Natl Acad Sci U S A* **85**:2815–9, 1988. OTE: (Medline: 88190159).

[Cheng-Mayer et al.(1990)] C. Cheng-Mayer, M. Quiroga, J. W. Tung, D. Dina, & J. A. Levy. Viral determinants of human immunodeficiency virus type 1 T-cell or macrophage tropism, cytopathogenicity, and CD4 antigen modulation. *J Virol* **64**:4390–8, 1990. OTE: (Medline: 90347835), (Genbank: M65024).

[Cheng-Mayer et al.(1991)] C. Cheng-Mayer, T. Shioda, & J. A. Levy. Host range, replicative, and cytopathic properties of human immunodeficiency virus type 1 are determined by very few amino acid changes in tat and gp120. *J Virol* **65**:6931–41, 1991. OTE: (Medline: 92046357).

[Ciesielski et al.(1992)] C. Ciesielski, D. Marianos, C. Y. Ou, R. Dumbaugh, J. Witte, R. Berkelman, B. Gooch, G. Myers, C. C. Luo, G. Schochetman, & et al. Transmission of human immunodeficiency virus in a dental practice. *Ann Intern Med* **116**:798–805, 1992. OTE: (Medline: 92231222).

[Ciesielski et al.(1994)] C. A. Ciesielski, D. W. Marianos, G. Schochetman, J. J. Witte, & H. W. Jaffe. The 1990 Florida dental investigation: The press and the science. *Ann Intern Med* **121**:886–888, 1994. OTE: (Medline: 95069397).

[Collman et al.(1992)] R. Collman, J. W. Balliet, S. A. Gregory, H. Friedman, D. L. Kolson, N. Nathanson, & A. Srinivasan. An infectious molecular clone of an unusual macrophage-tropic and highly cytopathic strain of human immunodeficiency virus type 1. *J Virol* **66**:7517–21, 1992. OTE: (Medline: 93059708), (Genbank: M96155).

[Cornelissen et al.(1995)] M. Cornelissen, G. Mulder-Kampinga, J. Veenstra, F. Zorgdrager, C. Kuiken, S. Hartman, J. Dekker, der van Hoek, C. Sol, & R. C. et al. Syncytium-inducing (SI) phenotype suppression at seroconversion after intramuscular inoculation of a non-syncytium-inducing/SI phenotypically mixed human immunodeficiency virus population. *J Virol* **69**:1810–1818, 1995. OTE: (Medline: 95156613), (Genbank: Z47411 Z47540).

[Couto-Fernandez et al.(1994)] J. C. Couto-Fernandez, W. Janssens, L. Heyndrickx, J. Motte, K. Fransen, M. Peeters, E. Delaporte, B. Galvao-Castro, P. Piot, & G. van der Groen. Genetic and antigenic variability of HIV type 1 in Brazil. *AIDS Res Hum Retroviruses* **10**:1157–1163, 1994. OTE: (Medline: 95127298).

[Crandall(1995)] K. A. Crandall. Intraspecific phylogenetics: support for dental transmission of human immunodeficiency virus. *J Virol* **69**:2351–2356, 1995. OTE: (Medline: 95191010).

[da Costa et al.(1995)] S. M. da Costa, M. Schechter, N. Shindo, A. C. Vicente, E. F. Oliveira, M. E. Pinto, & A. Tanuri. Sequence and phylogenetic analysis of glycoprotein 120 of an HIV type 1 variant (GWGR) prevalent in Brazil. *AIDS Res. Hum. Retroviruses* **11**:1143–5, 1995. OTE: (Medline: 96089224).

[Daniels et al.(1991)] R. S. Daniels, M. H. Smith, & A. G. Fisher. Molecular characterization of biologically diverse envelope variants of human immunodeficiency virus type 1 derived from an individual. *J Virol* **65**:5574–8, 1991. OTE: (Medline: 91374618).

[De Wolf et al.(1994)] F. De Wolf, E. Hogervorst, J. Goudsmit, E. M. Fenyo, H. Rubsamen-Waigmann, H. Holmes, B. Galvao-Castro, C. W. E. Karita, & S. D. Sempala. Syncytium-inducing and non-syncytium-inducing capacity of human immunodeficiency virus type 1 subtypes other than B: phenotypic and genotypic characteristics. WHO Network for HIV Isolation and Characterization. *AIDS Res Hum Retroviruses* **10**:1387–1400, 1994. OTE: (Medline: 95194697).

[Deacon et al.(1995)] N. J. Deacon, A. Tsykin, A. Solomon, K. Smith, M. Ludford-Menting, D. J. Hooker, D. A. McPhee, A. L. Greenway, A. Ellett, & C. C. et al. Genomic structure of an attenuated quasi

species of HIV-1 from a blood transfusion donor and recipients. *Science* **270**:988–991, 1995. OTE: (Medline: 96069819).

[DeBry et al.(1993)] R. W. DeBry, L. G. Abele, S. H. Weiss, M. D. Hill, M. Bouzas, E. Lorenzo, F. Graebnitz, & L. Resnick. Dental HIV transmission. *Nature* **361**:691, 1993. OTE: (Medline: 93180905).

[Delaporte et al.(1996)] E. Delaporte, W.Janssens, M. Peeters, A. Buve, G. Dibanga, J. Perret, V. Ditsambou, M. G. Courbot, A. Georges, A. Bourgeois, B. Samb, D. Henzel, L. Heyndrickx, K. Fransen, G. V. der Groen, B. Larouz, & J. Mbet. Epidemiological and molecular characteristics of hiv infection in gabon (1986 - 1994). *AIDS.* **10**:903–910, 1996.

[Delwart et al.(1995)] E. L. Delwart, M. P. Busch, M. L. Kalish, J. W. Mosley, & J. I. Mullins. Rapid molecular epidemiology of human immunodeficiency virus transmission. *AIDS Res. Hum. Retroviruses* **11**:1081–93, 1995. OTE: (Medline: 96089215).

[Delwart et al.(1994)] E. L. Delwart, H. W. Sheppard, B. D. Walker, J. Goudsmit, & J. I. Mullins. Human immunodeficiency virus type 1 evolution in vivo tracked by DNA heteroduplex mobility assays. *J Virol* **68**:6672–6683, 1994. OTE: (Medline: 94365971), (Genbank: U00821 U00822 U00832 U00833 U00834 U00837 U00839 U13240 U13241 U13242 U13243 U13244 U13245 U13246 U13247 U13248 U13249 U13250 U13251 U13252 U13373 U13374 U13375 U13376 U13377 U13378 U13379 U13380 U13381 U13382).

[Desai et al.(1986)] S. M. Desai, V. S. Kalyanaraman, J. M. Casey, A. Srinivasan, P. R. Andersen, & S. G. Devare. Molecular cloning and primary nucleotide sequence analysis of a distinct human immunodeficiency virus isolate reveal significant divergence in its genomic sequences. *Proc Natl Acad Sci U S A* **83**:8380–4, 1986. OTE: (Medline: 87041461), (Genbank: M13136 M13137).

[Diaz et al.(1995)] R. S. Diaz, E. C. Sabino, A. Mayer, J. W. Mosley, & M. P. Busch. Dual human immunodeficiency virus type 1 infection and recombination in a dually exposed transfusion recipient. The Transfusion Safety Study Group. *J Virol* **69**:3273–3281, 1995. OTE: (Medline: 95264414), (Genbank: U11124 U11125 U11126 U11127 U11128 U11129 U11130 U11131 U11132 U11133 U11134 U11135 U11136 U11137 U11138 U11139 U11140 U11141 U11142 U11143 U11144 U11145 U11146 U11147 U11148 U11149 U11150 U11151 U11152 U11153).

[Dietrich et al.(1993)] U. Dietrich, M. Grez, H. von Briesen, B. Panhans, M. Geissendorfer, H. Kuhnel, J. Maniar, G. Mahambre, W. B. Becker, & M. L. B. et al. HIV-1 strains from India are highly divergent from prototypic African and US/European strains, but are linked to a South African isolate. *AIDS* **7**:23–7, 1993. OTE: (Medline: 93183429), (Genbank: X65638 X65639 X65640 X68406 X68407 X65638 X65639 X65640 X68406 X68407).

[Douglas et al.(1996)] N. Douglas, A. Knight, A. Hayhurst, W. Barrett, M. Kevany, & R. Daniels. An efficient method for the rescue and analysis of functional hiv-1 env genes: evidence for recombination in the vicinity of the tat/rev splice site. *AIDS* **10**:39–46, 1996.

[Duensing et al.(1995)] T. D. Duensing, H. Fang, D. W. Dorward, & S. H. Pincus. Processing of the envelope glycoprotein gp160 in immunotoxin-resistant cell lines chronically infected with human immunodeficiency virus type 1. *J Virol* **69**:7122–7131, 1995. OTE: (Medline: 96013815).

[Dumitrescu et al.(1994)] O. Dumitrescu, M. L. Kalish, S. C. Kliks, C. I. Bandea, & J. A. Levy. Characterization of human immunodeficiency virus type 1 isolates from children in Romania: identification of a new envelope subtype. *J Infect Dis* **169**:281–8, 1994. OTE: (Medline: 94149311).

[Engelbrecht et al.(1995)] S. Engelbrecht, J. D. Laten, T. L. Smith, & E. J. van Rensburg. Identification of env subtypes in fourteen HIV type 1 isolates from south Africa. *AIDS Res. Hum. Retroviruses* **11**:1269–1271, 1995. OTE: (Medline: 96157218), (Genbank: U33769 U33770 U33771 U33772 U33773 U33774 U33775 U33776 U33777 U33778 U33779 U33780 U33781 U33782).

[Fisher et al.(1988)] A. G. Fisher, B. Ensoli, D. Looney, A. Rose, R. C. Gallo, M. S. Saag, G. M. Shaw, B. H. Hahn, & F. Wong-Staal. Biologically diverse molecular variants within a single HIV-1 isolate. *Nature* **334**:444–7, 1988. OTE: (Medline: 88302422).

[Fitzgibbon et al.(1993)] J. E. Fitzgibbon, S. Gaur, L. D. Frenkel, F. Laraque, B. R. Edlin, & D. T. Dubin. Transmission from one child to another of human immunodeficiency virus type 1 with a zidovudine-resistance mutation (see comments). *N Engl J Med* **329**:1835–41, 1993. OTE: (Medline: 94067210).

[Fouchier et al.(1992)] R. A. Fouchier, M. Groenink, N. A. Kootstra, M. Tersmette, H. G. Huisman, F. Miedema, & H. Schuitemaker. Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J Virol* **66**:3183–7, 1992. OTE: (Medline: 92219412).

[Furuta et al.(1994)] Y. Furuta, T. Bergstrom, G. Norkrans, & P. Horal. HIV type 1 V3 sequence diversity in contact-traced Swedish couples at the time of sexual transmission. *AIDS Res Hum Retroviruses* **10**:1187–1189, 1994. OTE: (Medline: 95127302), (Genbank: U10929 U10930 U10931 U10932 U10933 U10934 U10935 U10936 U10937 U10938 U10939 U10940 U10941 U10942 U10943 U10944 U10945 U10946 U10947 U10948 U10949 U10950).

[Gao et al.(1996a)] F. Gao, S. G. Morrison, D. L. Robertson, C. L. Thornton, S. Craig, G. Karlsson, J. Sodroski, M. Morgado, B. Galvao-Castro, H. von Briesen, S. Beddows, J. Weber, P. M. Sharp, G. M. Shaw, & B. H. Hahn. Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. The WHO and NIAID Networks for HIV Isolation and Characterization. *J. Virol.* **70**:1651–1657, 1996a. OTE: (Medline: 96190564), (Genbank: L34667 U02739 U04908 U08442 U08443 U08444 U08445 U08446 U08447 U08448 U08450 U08451 U08452 U08453 U08454 U08455 U08456 U08457 U08794 U08797 U08801 U09126 U09127 U09131 U27419 U27434 U27443 U43386).

[Gao et al.(1996b)] F. Gao, D. L. Robertson, S. G. S. G. Morrison, H. Hui, S. Craig, P. N. Fultz, J. Decker, M. Girard, G. M. Shaw, B. H. Hahn, & P. M. Sharp. The heterosexual HIV-1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J. Virology* **70**:7013–7029, 1996b.

[Gao et al.(1994a)] F. Gao, L. Yue, S. Craig, C. L. Thornton, D. L. Robertson, F. E. McCutchan, J. A. Bradac, P. M. Sharp, & B. H. Hahn. Genetic variation of HIV type 1 in four World Health Organization-sponsored vaccine evaluation sites: generation of functional envelope (glycoprotein 160) clones representative of sequence subtypes A, B, C, and E. WHO Network for HIV Isolation and Characterization. *AIDS Res Hum Retroviruses* **10**:1359–1368, 1994a. OTE: (Medline: 95194694), (Genbank: L34667 U08793 U08794 U08795 U08796 U08797 U08802 U08804 U08810 U09131).

[Gao et al.(1994b)] F. Gao, L. Yue, S. C. Hill, D. L. Robertson, A. H. Graves, M. S. Saag, G. M. Shaw, P. M. Sharp, & B. H. Hahn. HIV-1 sequence subtype D in the United States. *AIDS Res Hum Retroviruses* **10**:625–627, 1994b. OTE: (Medline: 95000937), (Genbank: U08192 U08193).

[Ghosh et al.(1993)] S. K. Ghosh, P. N. Fultz, E. Keddie, M. S. Saag, P. M. Sharp, B. H. Hahn, & G. M. Shaw. A molecular clone of HIV-1 tropic and cytopathic for human and chimpanzee lymphocytes. *Virology* **194**:858–64, 1993. OTE: (Medline: 93276579), (Genbank: L02317).

[Gomez et al.(1996)] C. E. Gomez, J. R. Fernandez, E. Iglesias, A. E. Lopez, L. Lobaina, E. Noa, H. Diaz, A. Herrera, F. Rolo, & C. A. Duarte. Complete DNA sequence of the gene encoding the external glycoprotein (gp120) from a Cuban HIV type 1 isolate. *AIDS Res. Hum. Retroviruses* **12**:553–555, 1996. OTE: (Medline: 96296968).

[Goodenow et al.(1989)] M. Goodenow, T. Huet, W. Saurin, S. Kwok, J. Sninsky, & S. Wain-Hobson. HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. *J Acquir Immune Defic Syndr* **2**:344–52, 1989. OTE: (Medline: 89328783).

[Grez et al.(1994)] M. Grez, U. Dietrich, P. Balfe, H. von Briesen, J. K. Maniar, G. Mahambre, E. L. Delwart, J. I. Mullins, & H. Rubsamen-Waigmann. Genetic analysis of human immunodeficiency virus type 1 and 2 (HIV-1 and HIV-2) mixed infections in India reveals a recent spread of HIV-1 and HIV-2 from a single ancestor for each of these viruses. *J Virol* **68**:2161–8, 1994. OTE: (Medline: 94187055).

[Guo et al.(1991)] H. G. Guo, J. C. Chermann, D. Waters, L. Hall, A. Louie, R. C. Gallo, H. Streicher, M. S. Reitz, M. Popovic, & W. Blattner. Sequence analysis of original HIV-1. *Nature* **349**:745–6, 1991. OTE: (Medline: 91156044), (Genbank: X57446 X57447 X57448 X57449 X57450 X57451 X57452 X57453 X57454 X57455).

[Gurgo et al.(1988)] C. Gurgo, H. G. Guo, G. Franchini, A. Aldovini, E. Collalti, K. Farrell, F. Wong-Staal, R. C. Gallo, & M. S. R. Jr. Envelope sequences of two new United States HIV-1 isolates. *Virology* **164**:531–6, 1988. OTE: (Medline: 88219542), (Genbank: M17449 M17450).

[Gurtler et al.(1994)] L. G. Gurtler, P. H. Hauser, J. Eberle, A. von Brunn, S. Knapp, L. Zekeng, J. M. Tsague, & L. Kaptue. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. *J Virol* **68**:1581–5, 1994. OTE: (Medline: 94149848).

[Hahn et al.(1986)] B. H. Hahn, G. M. Shaw, M. E. Taylor, R. R. Redfield, P. D. Markham, S. Z. Salahuddin, F. Wong-Staal, R. C. Gallo, E. S. Parks, & W. P. Parks. Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. *Science* **232**:1548–53, 1986. OTE: (Medline: 86235450), (Genbank: M12507).

[Hattori et al.(1991)] T. Hattori, K. Shiozaki, Y. Eda, S. Tokiyoshi, S. Matsushita, H. Inaba, M. Fujimaki, T. Meguro, K. Yamada, & M. H. et al. Characteristics of the principal neutralizing determinant of HIV-1 prevalent in Japan. *AIDS Res Hum Retroviruses* **7**:825–30, 1991. OTE: (Medline: 92075338), (Genbank: M81693 M81694 S70936).

[Heyndrickx et al.(1996)] L. Heyndrickx, W. Janssens, M. Alary, K. Fransen, K. Vereecken, S. Coppens, B. Willems, N. Davo, A. Guedeme, E. Baganizi, J. Joly, & G. van der Groen. Genetic variability of HIV type 1 in Benin. *AIDS Res. Hum. Retroviruses* **12**:1495–1497, 1996.

[Hillis & Huelsenbeck(1994)] D. M. Hillis & J. P. Huelsenbeck. Support for dental HIV transmission. *Nature* **369**:24–5, 1994. OTE: (Medline: 94217803), (Genbank: U06872 U06873 U06874 U06875 U06876 U06877 U06878 U06879 U06880 U06881 U06882 U06883 U06884 U06885 U06886 U06887 U06888 U06889 U06890 U06891 U06892 U06893 U06894 U06895 U06896 U06897 U06898 U06899 U06900 U06901).

[Ho et al.(1989)] D. D. Ho, H. S. Robin, M. Alam, B. J. Wallace, & Y. Mizrachi. Human immunodeficiency virus type 1 in a seronegative patient with visceral Kaposi's sarcoma and hypogammaglobulinemia. *American J of Med* **86**:349–51, 1989. OTE: (Medline: 89148307).

[Holm-Hansen et al.(1995)] C. Holm-Hansen, D. Grothues, S. Rustad, B. Rosok, F. R. Pascu, & B. Asjo. Characterization of HIV type 1 from Romanian children: lack of correlation between V3 loop amino acid sequence and syncytium formation in MT-2 cells. *AIDS Res Hum Retroviruses* **11**:597–603, 1995. OTE: (Medline: 96093895), (Genbank: X77968 X77969 X77970 X77971 X77972 X77973 X77974 X77975 X77976 X77977 X77978 X77979 X77980 X77981 X77982 X77983 X77984 X77985 X77986 X77987).

[Holmes et al.(1992)] E. C. Holmes, L. Q. Zhang, P. Simmonds, C. A. Ludlam, & A. J. Brown. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. *Proc Natl Acad Sci U S A* **89**:4835–9, 1992. OTE: (Medline: 92279221), (Genbank: M84240 M84241 M84242 M84243 M84244 M84245 M84246 M84247 M84248 M84249 M84250 M84251 M84252 M84253 M84254 M84255 M84256 M84257 M84258 M84259 M84260 M84261 M84262 M84263 M84264 M84265 M84266 M84267 M84268 M84269).

[Howard et al.(1994)] T. M. Howard, D. O. Olaylele, & S. Rasheed. Sequence analysis of the glycoprotein 120 coding region of a new HIV type 1 subtype A strain (HIV-1IbNg) from Nigeria. *AIDS Res Hum Retroviruses* **10**:1755–1757, 1994. OTE: (Medline: 95194745).

[Howard & Rasheed(1996)] T. M. Howard & S. Rasheed. Genomic structure and nucleotide sequence analysis of a new hiv type 1 subtype a strain from nigeria. *AIDS Res. Hum. Retroviruses* **12**:1413–1425, 1996.

[Huet et al.(1989)] T. Huet, M. C. Dazza, F. Brun-Vezinet, G. E. Roelants, & S. Wain-Hobson. A highly defective HIV-1 strain isolated from a healthy Gabonese individual presenting an atypical western blot. *AIDS* **3**:707–15, 1989. OTE: (Medline: 90148544).

[Hutto et al.(1996)] C. Hutto, Y. Zhou, J. He, R. Geffin, M. Hill, W. Scott, & C. Wood. Longitudinal studies of viral sequence, viral phenotype, and immunologic parameters of human immunodeficiency virus type 1 infection in perinatally infected twins with discordant disease courses. *J. Virol.* **70**:3589–3598, 1996. OTE: (Medline: 96211491).

[Hwang et al.(1991)] S. S. Hwang, T. J. Boyle, H. K. Lyerly, & B. R. Cullen. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science* **253**:71–4, 1991. OTE: (Medline: 91289160).

[Jaffe et al.(1994)] H. W. Jaffe, J. M. McCurdy, M. L. Kalish, T. Liberti, G. Metellus, B. H. Bowman, S. B. Richards, A. R. Neasman, & J. J. Witte. Lack of HIV transmission in the practice of a dentist with AIDS. *Ann Intern Med* **121**:855–859, 1994. OTE: (Medline: 95069391).

[Janssens et al.(1994a)] W. Janssens, L. Heyndrickx, de Van Peer, A. Bouckaert, K. Fransen, J. Motte, G. M. Gershy-Damet, M. Peeters, P. Piot, & G. van der Groen. Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cote d'Ivoire. *AIDS* **8**:21–6, 1994a. OTE: (Medline: 94280700).

[Janssens et al.(1994b)] W. Janssens, L. Heyndrickx, K. Fransen, J. Motte, M. Peeters, J. N. Nkengasong, P. M. Ndumbe, E. Delaporte, J. L. Perret, & C. A. et al. Genetic and phylogenetic analysis of env subtypes G and H in central Africa. *AIDS Res Hum Retroviruses* **10**:877–879, 1994b. OTE: (Medline: 95078007), (Genbank: U09664 U09665 U09666 U09667).

[Janssens et al.(1994c)] W. Janssens, L. Heyndrickx, K. Fransen, M. Temmerman, A. Leonaers, T. Ivens, J. Motte, P. Piot, & G. van der Groen. Genetic variability of HIV type 1 in Kenya. *AIDS Res Hum Retroviruses* **10**:1577–1579, 1994c. OTE: (Medline: 95194720), (Genbank: U12984 U12985 U12986 U12987 U12988 U12989 U12990 U12991 U12992 U12993 U12994 U12995 U12996 U12997 U12998 U12999 U13000 U13001 U13002 U13003 U13004 U13005 U13006).

[Janssens et al.(1994d)] W. Janssens, L. Heyndrickx, K. Fransen, M. Temmerman, A. Leonaers, T. Ivens, J. Motte, P. Piot, & der Van Groen. Genetic variability of HIV type 1 in Kenya. *AIDS Res Hum Retroviruses* **10**:1577–1579, 1994d. OTE: (Medline: 95194720), (Genbank: U12984 U12985 U12986 U12987 U12988 U12989 U12990 U12991 U12992 U12993 U12994 U12995 U12996 U12997 U12998 U12999 U13000 U13001 U13002 U13003 U13004 U13005 U13006).

[Janssens et al.(1994e)] W. Janssens, J. N. Nkengasong, L. Heyndrickx, K. Fransen, P. M. Ndumbe, E. Delaporte, M. Peeters, J. L. Perret, A. Ndoumou, & C. Atende. Further evidence of the presence of genetically aberrant HIV-1 strains in Cameroon and Gabon. *AIDS* **8**:1012–3, 1994e. OTE: (Medline: 95032906), (Genbank: X78476 X78477).

[Kaleebu et al.(1995)] P. Kaleebu, A. Bobkov, R. Cheingsong-Popov, P. Bieniasz, M. Garaev, & J. Weber. Identification of HIV-1 subtype G from Uganda. *AIDS Res Hum Retroviruses* **11**:657–659, 1995. OTE: (Medline: 96093903), (Genbank: U22010).

[Kalish et al.(1995)] M. L. Kalish, A. Baldwin, S. Raktham, C. Wasi, C. C. Luo, G. Schochetman, T. D. Mastro, N. Young, S. Vanichseni, & H. R. et al. The evolving molecular epidemiology of HIV-1 envelope subtypes in injecting drug users in Bangkok, Thailand: implications for HIV vaccine trials. *AIDS* **9**:851–857, 1995. OTE: (Medline: 96014957).

[Kalish et al.(1994)] M. L. Kalish, C. C. Luo, B. G. Weniger, K. Limpakarnjanarat, N. Young, C. Y. Ou, & G. Schochetman. Early HIV type 1 strains in Thailand were not responsible for the current epidemic. *AIDS Res Hum Retroviruses* **10**:1573–1575, 1994. OTE: (Medline: 95194719), (Genbank: L15724 L15728 L15730 L19237 L19238 L19239 L19571 L19579 L19634 U15576 U15577 U15578 U15579 U15580 U15581 U15582 U15583 U15584 U15585 U15586 U15587 U15588).

[Kasper et al.(1994)] P. Kasper, R. Kaiser, J. Oldenburg, H. H. Brackmann, B. Matz, & K. E. Schneweis. Parallel evolution in the V3 region of HIV type 1 after infection of hemophiliacs from a homogeneous source. *AIDS Res Hum Retroviruses* **10**:1669–1678, 1994. OTE: (Medline: 95194733).

[Keys et al.(1993)] B. Keys, J. Karis, B. Fadeel, A. Valentin, G. Norkrans, L. Hagberg, & F. Chiodi. V3 sequences of paired HIV-1 isolates from blood and cerebrospinal fluid cluster according to host and show variation related to the clinical stage of disease. *Virology* **196**:475–83, 1993. OTE: (Medline: 93383372), (Genbank: Z23177 Z23178 Z23179 Z23180 Z23181 Z23182 Z23183 Z23184 Z23185 Z23186 Z23187 Z23188 Z23190 Z23191 Z23192 Z23193 Z23194 Z23195 Z23196 Z23197 Z23198 Z23199 Z23200 Z23201 Z23202 Z23203 Z23204 Z23205 Z23206 Z23207).

[Kim et al.(1995)] F. M. Kim, D. L. Kolson, J. W. Balliet, A. Srinivasan, & R. G. Collman. V3-independent determinants of macrophage tropism in a primary human immunodeficiency virus type 1 isolate. *J Virol* **69**:1755–1761, 1995. OTE: (Medline: 95156606).

[Klasse et al.(1996)] P. J. Klasse, M. T. Boyd, R. A. Weiss, & T. F. Schulz. Mutations in the vpu, env and nef genes of a syncytium-inducing variant of HIV type 1 JR-CSF that infects a range of T-cell lines. *AIDS Res. Hum. Retroviruses* **12**:347–350, 1996.

[Komiyama et al.(1989)] N. Komiyama, N. Hattori, J. Inoue, S. Sakuma, T. Kurimura, & M. Yoshida. Nucleotide sequences of gag and env genes of a Japanese isolate of HIV- 1 and their expression in bacteria. *AIDS Res Hum Retroviruses* **5**:411–9, 1989. OTE: (Medline: 89352108).

[Korber & Myers(1992)] B. Korber & G. Myers. Signature pattern analysis: a method for assessing viral sequence relatedness. *AIDS Res Hum Retroviruses* **8**:1549–60, 1992. OTE: (Medline: 93090472), (Genbank: M90847 M90848 M90849 M90850 M90851 M90852 M90853 M90854 M90855 M90856 M90857 M90858 M90859 M90860 M90861 M90862 M90863 M90864 M90865 M90866 M90867 M90868 M90869 M90870 M90871 M90872 M90873 M90874 M90875 M90876).

[Korber et al.(1994)] B. T. Korber, K. J. Kunstman, B. K. Patterson, M. Furtado, M. M. McEvilly, R. Levy, & S. M. Wolinsky. Genetic differences between blood- and brain-derived viral sequences from human immunodeficiency virus type 1-infected patients: evidence of conserved elements in the V3 region of the envelope protein of brain- derived sequences. *J Virol* **68**:7467–7481, 1994. OTE: (Medline: 95018660), (Genbank: U05360 U05361 U05362 U05363 U05364 U05365 U05366 U05367 U05368 U05369 U05370 U05371 U05372 U05373 U05374 U05375 U05376 U05377 U05378 U05379 U05380 U05381 U05382 U05383 U05384 U05385 U05386 U05387 U05388 U05389).

[Korber et al.(1995)] B. T. Korber, G. Learn, J. I. Mullins, B. H. Hahn, & S. Wolinsky. Protecting HIV databases. *Nature* **378**:242–244, 1995. OTE: (Medline: 96069755).

[Kostrikis et al.(1995)] L. G. Kostrikis, E. Bagdades, Y. Cao, L. Zhang, D. Dimitriou, & D. D. Ho. Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: identification of a new subtype designated subtype I. *J Virol* **69**:6122–6130, 1995. OTE: (Medline: 95395946), (Genbank: L19573 L19576 L22084 L22954 L22955 M15896 M38428 M66533 U28321 U28661 U28662 U28663 U28664 U28665 U28666 U28667 U28668 U28669 U28670 U28671 U28672 U28673 U28674 U28675 U28676 U28677 U28678 U28679 U28680 U28719).

[Kreutz et al.(1992)] R. Kreutz, U. Dietrich, H. Kuhnel, K. Nieselt-Struwe, M. Eigen, & H. Rubsamen-Waigmann. Analysis of the envelope region of the highly divergent HIV-2ALT isolate extends the known range of variability within the primate immunodeficiency viruses. *AIDS Res Hum Retroviruses* **8**:1619–29, 1992. OTE: (Medline: 93090480), (Genbank: X61240 X61240).

[Kuiken et al.(1996a)] C. Kuiken, V. Lukashov, E. Baan, J. dekker, J. Leunissen, & J. Goudsmit. Evidence for limited intra-subject evolution of the v3 domain of the hiv-1 envelope in the amsterdam population. *AIDS.* **In Press**, 1996a.

[Kuiken et al.(1996b)] C. L. Kuiken, M. T. Cornelissen, F. Zorgdrager, S. Hartman, A. J. Gibbs, & J. Goudsmit. Consistent risk group-associated differences in human immunodeficiency virus type 1 vpr, vpu and V3 sequences despite independent evolution. *J. Gen. Virol.* **77**:783–792, 1996b. OTE: (Medline: 96203966), (Genbank: Z29280 Z29296 Z29298 Z29299 Z29302 Z29303 Z29304 Z29306 Z29307 Z29310 Z29311 Z29312 Z29315 Z29319 Z29320 Z29321 Z29322 Z29323 Z29324 Z68033 Z68047 Z68062 Z68070 Z68074 Z68086 Z68508 Z68509 Z68510 Z68511 Z68512).

[Kuiken et al.(1995)] C. L. Kuiken, J. Goudsmit, G. F. Weiller, J. S. Armstrong, S. Hartman, P. Portegies, J. Dekker, & M. Cornelissen. Differences in human immunodeficiency virus type 1 V3 sequences from patients with and without AIDS dementia complex. *J Gen Virol* **76**:175–180, 1995. OTE: (Medline: 95146947), (Genbank: Z37734 Z37735 Z37736 Z37737 Z37738 Z37739 Z37740 Z37741 Z37742 Z37743 Z37744 Z37745 Z37746 Z37747 Z37748 Z37749 Z37750 Z37751 Z37752 Z37753 Z37754 Z37755 Z37756 Z37757 Z37758 Z37759 Z37760 Z37761 Z37762 Z37763).

[Kuiken et al.(1993)] C. L. Kuiken, G. Zwart, E. Baan, R. A. Coutinho, J. A. van den Hoek, & J. Goudsmit. Increasing antigenic and genetic diversity of the V3 variable domain of the human immunodeficiency virus envelope protein in the course of the AIDS epidemic. *Proc Natl Acad Sci U S A* **90**:9061–5, 1993. OTE: (Medline: 94022315).

[Kusumi et al.(1992)] K. Kusumi, B. Conway, S. Cunningham, A. Berson, C. Evans, A. K. Iversen, D. Colvin, M. V. Gallo, S. Coutre, & E. G. S. et al. Human immunodeficiency virus type 1 envelope gene structure and diversity in vivo and after cocultivation in vitro. *J Virol* **66**:875–85, 1992. OTE: (Medline: 92114191), (Genbank: M79342 M79343 M79344 M79345 M79346 M79347 M79348 M79349 M79350 M79351 M79352 M79353 M79354).

[Learn et al.(1996)] G. H. Learn, B. T. M. Korber, B. Foley, B. H. Hahn, S. M. Wolinsky, & J. I. Mullins. Maintaining the integrity of human immunodeficiency virus sequence databases. **70**:5720–30, 1996.

[Leitner et al.(1995)] T. Leitner, A. Alaeus, S. Marquina, E. Lilja, K. Lidman, & J. Albert. Yet another subtype of HIV type 1. *AIDS Res Hum Retroviruses* **11**:995–997, 1995. OTE: (Medline: 96020102), (Genbank: L41176 L41177 L41178 L41179).

[R. et al.(1990)] L. D. R., B. Vanderborght, M. Vanden Haesevelde, L. Heyndrickx, A. van Geel, C. Wauters, R. Bernaerts, E. Saman, P. Nijs, & B. W. et al. Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. *J Virol* **64**:1207–16, 1990. OTE: (Medline: 90156518), (Genbank: M31171).

[T. Leitner et al.(1996)] T. T. Leitner, D. D. Escanilla, C. Franzen, M. Uhlen, & J. Albert. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proc. Natl. Acad. Sci. U.S.A.* **20**:10864–10869, 1996.

[et al(1992)] T. F. W. W. et al. Genetic variation of the third variable envelope region in natural HIV-1 infection. *Doctoral Thesis, Amsterdam* 1992.

[Jr et al.(1992)] M. S. R. Jr, H. G. Guo, J. Oleske, J. Hoxie, M. Popovic, E. Read-Connole, P. Markham, H. Streicher, & R. C. Gallo. On the historical origins of HIV-1 (MN) and (RF). *AIDS Res Hum Retroviruses* **8**:1950, 1992. OTE: (Medline: 93144005).

[Jr et al.(1994)] M. S. R. Jr, L. Hall, M. Robert-Guroff, J. Lautenberger, B. M. Hahn, G. M. Shaw, L. I. Kong, S. H. Weiss, D. Waters, & R. C. G. et al. Viral variability and serum antibody response in a laboratory worker infected with HIV type 1 (HTLV type IIIB). *AIDS Res Hum Retroviruses* **10**:1143–1155, 1994. OTE: (Medline: 95127297).

[Levy et al.(1984)] J. A. Levy, A. D. Hoffman, S. M. Kramer, J. A. Landis, J. M. Shimabukuro, & L. S. Oshiro. Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science* **225**:840–2, 1984. OTE: (Medline: 84300260).

[Li et al.(1992)] Y. Li, H. Hui, C. J. Burgess, R. W. Price, P. M. Sharp, B. H. Hahn, & G. M. Shaw. Complete nucleotide sequence, genome organization, and biological properties of human immunodeficiency virus type 1 in vivo: evidence for limited defectiveness and complementation. *J Virol* **66**:6587–600, 1992. OTE: (Medline: 93021387), (Genbank: M93258 M93259).

[Li et al.(1991)] Y. Li, J. C. Kappes, J. A. Conway, R. W. Price, G. M. Shaw, & B. H. Hahn. Molecular characterization of human immunodeficiency virus type 1 cloned directly from uncultured human brain tissue: identification of replication-competent and -defective viral genomes. *J Virol* **65**:3973–85, 1991. OTE: (Medline: 91303644).

[Liu et al.(1990)] Z. Q. Liu, C. Wood, J. A. Levy, & C. Cheng-Mayer. The viral envelope gene is involved in macrophage tropism of a human immunodeficiency virus type 1 strain isolated from brain tissue. *J Virol* **64**:6148–53, 1990. OTE: (Medline: 91056585), (Genbank: M38673 M38673).

[Loussert-Ajaka et al.(1995)] I. Loussert-Ajaka, M. L. Chaix, B. Korber, F. Letourneur, E. Gomas, E. Allen, T. D. Ly, F. Brun-Vezinet, F. Simon, & S. Saragosti. Variability of human immunodeficiency virus type 1 group O strains isolated from Cameroonian patients living in France. *J Virol* **69**:5640–5649, 1995. OTE: (Medline: 95363977), (Genbank: U24562 U24563 U24564 U24565 U24566 U24567 U24568 U24706 U24707 U24708 U24709 U24710 U24711 U24712).

[Louwagie et al.(1994)] J. Louwagie, E. L. Delwart, J. I. Mullins, F. E. McCutchan, G. Eddy, & D. S. Burke. Genetic analysis of HIV-1 isolates from Brazil reveals presence of two distinct genetic subtypes. *AIDS Res Hum Retroviruses* **10**:561–567, 1994. OTE: (Medline: 95000928), (Genbank: L11751 L11752 L11753 L11754 L22082 L22083 L22084 L22085 L22086 L22087 L22088).

[Louwagie et al.(1995)] J. Louwagie, W. Janssens, J. Mascola, L. Heyndrickx, P. Hegerich, der van Groen, F. E. McCutchan, & D. S. Burke. Genetic diversity of the envelope glycoprotein from human immunodeficiency virus type 1 isolates of African origin. *J Virol* **69**:263–271, 1995. OTE: (Medline: 95074874), (Genbank: L22939 L22957 L23064 L23065).

[Louwagie et al.(1993)] J. Louwagie, F. E. McCutchan, M. Peeters, T. P. Brennan, E. Sanders-Buell, G. A. Eddy, G. van der Groen, K. Fransen, G. M. Gershy-Damet, & R. D. et al. Phylogenetic analysis of gag genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. *AIDS* **7**:769–80, 1993. OTE: (Medline: 93371703), (Genbank: L03696 L03702 L03705 L03707 L11751 L11752 L11753 L11754 L11755 L11756 L11757 L11758 L11759 L11760 L11761 L11762 L11763 L11764 L11765 L11766 L11767 L11768 L11769 L11770 L11771 L11772 L11773 L11774 L11775 L11776).

[Lukashov et al.(1996)] V. Lukashov, C. Kuiken, K. Boer, & J. Goudsmit. Hiv type 1 subtypes in the netherlands circulating among women originating from aids-endemic regions. *AIDS Res. Hum. Retroviruses* **12**:951–953, 1996.

[Lukashov et al.(1995)] V. V. Lukashov, M. T. Cornelissen, J. Goudsmit, M. N. Papuashvilli, P. G. Rytik, R. M. Khaitov, E. V. Karamov, & W. de F. Simultaneous introduction of distinct HIV-1 subtypes into different risk groups in Russia, Byelorussia and Lithuania. *AIDS* **9**:435–439, 1995. OTE: (Medline: 95367213).

[Lukashov & Goudsmit(1995)] V. V. Lukashov & J. Goudsmit. Increasing genotypic and phenotypic selection from the original genomic RNA populations of HIV-1 strains LAI and MN (NM) by peripheral blood mononuclear cell culture, B-cell-line propagation and T-cell-line adaptation. *AIDS* **9**:1307–1311, 1995. OTE: (Medline: 96188304).

[Marquina et al.(1996)] S. Marquina, T. Leitner, R. D. Rabinovich, J. Benetucci, O. Libonatti, & J. Albert. Co-existence of subtypes B, F and a B/F env recombinant of HIV-1 in Buenos Aires, Argentina. *AIDS Res. Hum. Retroviruses* **12**:1651–1654, 1996.

[Mascola et al.(1994)] J. R. Mascola, J. Louwagie, F. E. McCutchan, C. L. Fischer, P. A. Hegerich, K. F. Wagner, A. K. Fowler, J. G. McNeil, & D. S. Burke. Two antigenically distinct subtypes of human immunodeficiency virus type 1: viral genotype predicts neutralization serotype. *J Infect Dis* **169**:48–54, 1994. OTE: (Medline: 94103665), (Genbank: L14570 L14571 L14572 L14573 L14574 L14575 L14576).

[McCutchan et al.(1996)] F. E. McCutchan, A. W. Artenstein, E. Sanders-Buell, M. O. Salminen, J. K. Carr, J. R. Mascola, X. F. Yu, K. E. Nelson, C. Khamboonruang, D. Schmitt, M. P. Kieny, J. G. McNeil, & D. S. Burke. Diversity of the envelope glycoprotein among human immunodeficiency virus type 1 isolates of clade E from Asia and Africa. *J. Virol.* **70**:3331–3338, 1996. OTE: (Medline: 96211461), (Genbank: U48264 U48265 U48266 U48267 U48268 U48269 U48272 U48723 U48724 U48725 U48726 U48727 U48728).

[McCutchan et al.(1992)] F. E. McCutchan, P. A. Hegerich, T. P. Brennan, P. Phanuphak, P. Singharaj, A. Jugsudee, P. W. Berman, A. M. Gray, A. K. Fowler, & D. S. Burke. Genetic variants of HIV-1 in Thailand. *AIDS Res Hum Retroviruses* **8**:1887–95, 1992. OTE: (Medline: 93143998).

[McKeating et al.(1996)] J. A. McKeating, Y. J. Zhang, C. Arnold, R. Frederiksson, E. M. Fenyo, & P. Balfe. Chimeric viruses expressing primary envelope glycoproteins of human immunodeficiency virus type I show increased sensitivity to neutralization by human sera. *Virology* **220**:450–460, 1996. OTE: (Medline: 96251941), (Genbank: U57788 U57789 U57790 U57791 U57792 U57793 U57794).

[McNearney et al.(1993)] T. McNearney, Z. Hornickova, B. Kloster, A. Birdwell, G. A. Storch, S. H. Polmar, M. Arens, & L. Ratner. Evolution of sequence divergence among human immunodeficiency virus type 1 isolates derived from a blood donor and a recipient. *Pediatr Res* **33**:36–42, 1993. OTE: (Medline: 93165379).

[McNearney et al.(1992)] T. McNearney, Z. Hornickova, R. Markham, A. Birdwell, M. Arens, A. Saah, & L. Ratner. Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. *Proc Natl Acad Sci U S A* **89**:10247–51, 1992. OTE: (Medline: 93066216).

[McNearney et al.(1990)] T. McNearney, P. Westervelt, B. J. Thielan, D. B. Trowbridge, J. Garcia, R. Whittier, & L. Ratner. Limited sequence heterogeneity among biologically distinct human immunodeficiency virus type 1 isolates from individuals involved in a clustered infectious outbreak. *Proc Natl Acad Sci U S A* **87**:1917–21, 1990. OTE: (Medline: 90175407), (Genbank: M31451).

[Menu et al.(1996)] E. Menu, T.X., Truong, M. Lafon, T. Nguyen, M. Muller-Trutwin, T. Nguyen, A. Deslandres, G. Chaouat, Q. Duong, B. Ha, H. Fleury, & F. Barre-Sinoussi. Hiv type 1 thai subtype e is predominant in south vietnam. *AIDS Res. Hum. Retroviruses* **12**:629–633, 1996.

[Monken et al.(1995)] C. E. Monken, B. Wu, & A. Srinivasan. High resolution analysis of HIV-1 quasispecies in the brain. *AIDS* **9**:345–349, 1995. OTE: (Medline: 95314789), (Genbank: L17088 L17089 L17090 L17091 L17092 L17093 L17094 L17095 L17096 L17097 L17098 L17099 L17100 L17101 L17102 L17103 L17104 L17105 L17106 L17107 L17108 L17109 L17110 L17111 L17112 L17113 L17114 L17115 L17116 L17117).

[Montpetit(1995)] M. Montpetit. HIV-1 subtype A in Canada. *AIDS Res. Hum. Retroviruses* **11**:1421–1422, 1995. OTE: (Medline: 96159142).

[Morgado et al.(1994)] M. G. Morgado, E. C. Sabino, E. G. Shpaer, V. Bongertz, L. Brigido, M. D. Guimaraes, E. A. Castilho, B. Galvao-Castro, J. I. Mullins, & R. M. H. et al. V3 region polymorphisms in HIV-1 from Brazil: prevalence of subtype B strains divergent from North American/European prototype and detection of subtype F. *AIDS Res Hum Retroviruses* **10**:569–576, 1994. OTE: (Medline: 95000929), (Genbank: L19237 L19571 L19575 L19576 L19579).

[Mulder-Kampinga et al.(1993)] G. A. Mulder-Kampinga, C. Kuiken, J. Dekker, H. J. Scherpbier, K. Boer, & J. Goudsmit. Genomic human immunodeficiency virus type 1 RNA variation in mother and child following intra-uterine virus transmission. *J Gen Virol* **74**:1747–56, 1993. OTE: (Medline: 93389428), (Genbank: L21028 L21153).

[Mulder-Kampinga et al.(1995)] G. A. Mulder-Kampinga, A. Simonon, C. L. Kuiken, J. Dekker, H. J. Scherpbier, de van Perre, K. Boer, & J. Goudsmit. Similarity in env and gag genes between genomic RNAs of human immunodeficiency virus type 1 (HIV-1) from mother and infant is unrelated to time of HIV-1 RNA positivity in the child. *J Virol* **69**:2285–2296, 1995. OTE: (Medline: 95191002), (Genbank: L21028 L21029 L21030 L21031 L21032 L21033 L21034 L21035 L21036 L21037 L21038 L21039 L21040 L21041 L21042 L21043 L21044 L21045 L21046 L21047 L21048 L21049 L21050 L21051 L21052 L21053 Z47817 Z47818 Z47819 Z47820).

[Murphy et al.(1993)] E. Murphy, B. Korber, M. C. Georges-Courbot, B. You, A. Pinter, D. Cook, M. P. Kieny, A. Georges, C. Mathiot, & F. B. et al. Diversity of V3 region sequences of human immunodeficiency viruses type 1 from the central African Republic. *AIDS Res Hum Retroviruses* **9**:997–1006, 1993. OTE: (Medline: 94107601), (Genbank: L11457 L11458 L11459 L11460 L11461 L11462

L11463 L11464 L11465 L11466 L11467 L11468 L11469 L11470 L11471 L11472 L11473 L11474 L11475 L11476 L11477 L11478 L11479 L11480 L11481 L11482 L11483 L11484 L11485 L11486).

[Nerurkar et al.(1996)] V. R. Nerurkar, H. T. Nguyen, W. M. Dashwood, P. R. Hoffmann, C. Yin, D. M. Morens, A. H. Kaplan, R. Detels, & R. Yanagihara. HIV type 1 subtype E in commercial sex workers and injection drug users in Southern Vietnam. *AIDS Res. Hum. Retroviruses* **12**:841–843, 1996.

[Nkengasong et al.(1994)] J. N. Nkengasong, W. Janssens, L. Heyndrickx, K. Fransen, P. M. Ndumbe, J. Motte, A. Leonaers, M. Ngolle, J. Ayuk, & P. P. et al. Genotypic subtypes of HIV-1 in Cameroon. *AIDS* **8**:1405–1412, 1994. OTE: (Medline: 95118531).

[O'Brien et al.(1990)] W. A. O'Brien, Y. Koyanagi, A. Namazie, J. Q. Zhao, A. Diagne, K. Idler, J. A. Zack, & I. S. Chen. HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120 outside the CD4-binding domain. *Nature* **348**:69–73, 1990. OTE: (Medline: 91043044).

[Oram et al.(1990)] J. D. Oram, R. G. Downing, M. Roff, J. C. Clegg, D. Serwadda, & J. W. Carswell. Nucleotide sequence of a Ugandan HIV-1 provirus reveals genetic diversity from other HIV-1 isolates. *AIDS Res Hum Retroviruses* **6**:1073–8, 1990. OTE: (Medline: 91090981), (Genbank: M62320).

[Oram et al.(1991)] J. D. Oram, R. G. Downing, M. Roff, N. Serwankambo, J. C. Clegg, A. S. Featherstone, & J. C. Booth. Sequence analysis of the V3 loop regions of the env genes of Ugandan human immunodeficiency proviruses. *AIDS Res Hum Retroviruses* **7**:605–14, 1991. OTE: (Medline: 92118457).

[Orloff et al.(1993)] G. M. Orloff, M. L. Kalish, J. Chiphangwi, K. E. Potts, C. Y. Ou, G. Schochetman, G. Dallabetta, A. I. Saah, & P. G. Miotti. V3 loops of HIV-1 specimens from pregnant women in Malawi uniformly lack a potential N-linked glycosylation site. *AIDS Res Hum Retroviruses* **9**:705–6, 1993. OTE: (Medline: 93378795), (Genbank: L15721 L15722 L15723 L15724 L15725 L15726 L15727 L15728 L15729 L15730 L15731 L15732 L15733 L15734 L15735).

[Osmanov et al.(1994)] S. Osmanov, W. L. Heyward, & J. Esparza. The World Health Organization Network for HIV Isolation and Characterization: summary of a pilot study. *AIDS Res Hum Retroviruses* **10**:1325–1326, 1994. OTE: (Medline: 95194690).

[Ou et al.(1992a)] C. Y. Ou, C. A. Ciesielski, G. Myers, C. I. Bandea, C. C. Luo, B. T. Korber, J. I. Mullins, G. Schochetman, R. L. Berkelman, & A. N. E. et al. Molecular epidemiology of HIV transmission in a dental practice. *Science* **256**:1165–71, 1992a. OTE: (Medline: 92271245), (Genbank: M90847 M90848 M90849 M90850 M90851 M90852 M90853 M90854 M90855 M90856 M90857 M90858 M90859 M90860 M90861 M90862 M90863 M90864 M90865 M90866 M90867 M90868 M90869 M90870 M90871 M90872 M90873 M90874 M90875 M90876).

[Ou et al.(1992b)] C. Y. Ou, Y. Takebe, C. C. Luo, M. Kalish, W. Auwanit, C. Bandea, la de Torre, J. L. Moore, G. Schochetman, & S. Y. et al. Wide distribution of two subtypes of HIV-1 in Thailand. *AIDS Res Hum Retroviruses* **8**:1471–2, 1992b. OTE: (Medline: 93103845).

[Ou et al.(1993)] C. Y. Ou, Y. Takebe, B. G. Weniger, C. C. Luo, M. L. Kalish, W. Auwanit, S. Yamazaki, H. D. Gayle, N. L. Young, & G. Schochetman. Independent introduction of two major HIV-1 genotypes into distinct high-risk populations in Thailand. *Lancet* **341**:1171–4, 1993. OTE: (Medline: 93254133).

[Palca(1992a)] J. Palca. AIDS. CDC closes the case of the Florida dentist. *Science* **256**:1130–1, 1992a. OTE: (Medline: 92271233).

[Palca(1992b)] J. Palca. The case of the Florida dentist. *Science* **255**:392–4, 1992b. OTE: (Medline: 92132539).

[Pang et al.(1991)] S. Pang, H. V. Vinters, T. Akashi, W. A. O'Brien, & I. S. Chen. HIV-1 env sequence variation in brain tissue of patients with AIDS- related neurologic disease. *J Acquir Immune Defic Syndr* **4**:1082–92, 1991. OTE: (Medline: 92092169).

[Penny et al.(1996)] M. A. Penny, S. J. Thomas, N. W. Douglas, S. Ranjbar, H. Holmes, & R. S. Daniels. Env-gene sequences of primary HIV-1 isolates of subtypes B, C, D, E and F obtained from the WHO Network for HIV Isolation and Characterisation. *AIDS Res Hum Retroviruses* **12**:741–747, 1996.

[Pestano et al.(1993)] G. Pestano, A. Prince, J. Guyden, J. M. Ntambi, A. Atkin, & W. M. Boto. Independent divergences in the CD4 binding site and V3 loop encoded in two seroprevalent Ugandan HIV-1 clinical isolates. *J Acquir Immune Defic Syndr* **6**:872–80, 1993. OTE: (Medline: 93301830), (Genbank: M98503 M98504).

[Pestano et al.(1995)] G. A. Pestano, K. S. Hosford, A. I. Spira, J. Riley, J. M. Xie, N. Sewankambo, L. Brown, D. D. Ho, & W. M. Boto. Seroreactivity of analogous antigenic epitopes in glycoprotein 120 expressed in HIV-1 subtypes A, B, C, and D. *AIDS Res Hum Retroviruses* **11**:589–596, 1995. OTE: (Medline: 96093894).

[Pincus et al.(1994)] S. H. Pincus, K. G. Messer, P. L. Nara, W. A. Blattner, G. Colclough, & M. Reitz. Temporal analysis of the antibody response to HIV envelope protein in HIV-infected laboratory workers. *J Clin Invest* **93**:2505–13, 1994. OTE: (Medline: 94259810).

[Popovic et al.(1984)] M. Popovic, N. Flomenberg, D. J. Volkman, D. Mann, A. S. Fauci, B. Dupont, & R. C. Gallo. Alteration of T-cell functions by infection with HTLV-I or HTLV-II. *Science* **226**:459–62, 1984. OTE: (Medline: 85040344).

[Potts et al.(1993a)] K. E. Potts, M. L. Kalish, C. I. Bandea, G. M. Orloff, L. S. M., C. Brown, N. Malanda, M. Kavuka, G. Schochetman, & C. Y. O. et al. Genetic diversity of human immunodeficiency virus type 1 strains in Kinshasa, Zaire. *AIDS Res Hum Retroviruses* **9**:613–8, 1993a. OTE: (Medline: 93378782).

[Potts et al.(1993b)] K. E. Potts, M. L. Kalish, T. Lott, G. Orloff, C. C. Luo, M. A. Bernard, C. B. Alves, R. Badaro, J. Suleiman, & O. F. et al. Genetic heterogeneity of the V3 region of the HIV-1 envelope glycoprotein in Brazil. Brazilian Collaborative AIDS Research Group. *AIDS* **7**:1191–7, 1993b. OTE: (Medline: 94030749).

[Quinones-Mateu et al.(1995)] M. E. Quinones-Mateu, J. Dopazo, J. A. Este, T. R. Rota, & E. Domingo. Molecular characterization of human immunodeficiency virus type 1 isolates from Venezuela. *AIDS Res Hum Retroviruses* **11**:605–616, 1995. OTE: (Medline: 96093896), (Genbank: U16764 U16779).

[Quinones-Mateu et al.(1996)] M. E. Quinones-Mateu, A. Holguin, V. Soriano, & E. Domingo. Env gene diversity of HIV type 1 isolates from Spain. *AIDS Res. Hum. Retroviruses* **12**:955–957, 1996.

[Ranjbar(1995)] S. Ranjbar. Molecular characterization of an HIV type 1 isolate from Burundi. *AIDS Res. Hum. Retroviruses* **11**:1146, 1995. OTE: (Medline: 96089225).

[Ranjbar et al.(1995)] S. Ranjbar, A. Slade, A. Jenkins, A. Heath, P. Kitchin, N. Almond, S. Osmanov, & H. Holmes. Molecular characterization of an HIV type 1 isolate from Burundi. *AIDS Res Hum Retroviruses* **11**:981–984, 1995. OTE: (Medline: 96020099), (Genbank: L35452 L35453 L35454 L35455 L35456 L35457 L35458 L35459).

[Richalet-Secordel et al.(1994)] P. M. Richalet-Secordel, A. Deslandres, S. Plaue, B. You, F. Barre-Sinoussi, & R. V. M. H. Cross-reactive potential of rabbit antibodies raised against a cyclic peptide representing a chimeric V3 loop of HIV-1 gp120 studied by biosensor technique and ELISA. *FEMS Immunol Med Microbiol* **9**:77–87, 1994. OTE: (Medline: 95004283).

[Robbins et al.(1996)] K. E. Robbins, C. I. Bandea, A. Levin, J. J. Goedert, W. A. Blattner, G. Brubaker, T. M. Brown, G. Schochetman, M. L. Kalish, J. Shao, & T. R. O'Brien. Genetic variability of human immunodeficiency virus type 1 in rural northwest Tanzania. *AIDS Res. Hum. Retroviruses* **12**:1389–1391, 1996.

[Roth et al.(1996)] W. W. Roth, J. A. Zuberi, H. G. Stringer, S. K. Davidson, & V. C. Bond. Examination of HIV type 1 variants in mother-child pairs. *AIDS Res. Hum. Retroviruses* **12**:925–930, 1996.

[Sabino et al.(1994a)] E. Sabino, L. Z. Pan, C. Cheng-Mayer, & A. Mayer. Comparison of in vivo plasma and peripheral blood mononuclear cell HIV- 1 quasi-species to short-term tissue culture isolates: an analysis of tat and C2-V3 env regions. *AIDS* **8**:901–9, 1994a. OTE: (Medline: 95032920), (Genbank: U01357 U01358 U01359 U01360 U01361 U01362 U01363 U01364 U01365 U01366 U01367 U01368 U01369 U01370 U01371 U01372 U01373 U01374 U01375 U01379 U01380 U01381 U01382 U01383 U01384 U01385 U01386 U01387 U01388 U01389).

[Sabino et al.(1994b)] E. C. Sabino, E. Delwart, T. H. Lee, A. Mayer, J. I. Mullins, & M. P. Busch. Identification of low-level contamination of blood as basis for detection of human immunodeficiency virus (HIV) DNA in anti-HIV- negative specimens. *J Acquir Immune Defic Syndr* **7**:853–859, 1994b. OTE: (Medline: 94293167), (Genbank: L20371 L20372 L20373 L20374 L20375 L20376 L20377 L20378 L20379 L20380).

[Sabino et al.(1994c)] E. C. Sabino, E. G. Shpaer, M. G. Morgado, B. T. Korber, R. S. Diaz, V. Bongertz, S. Cavalcante, B. Galvao-Castro, J. I. Mullins, & A. Mayer. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. *J Virol* **68**:6340–6346, 1994c. OTE: (Medline: 94365938), (Genbank: U08955 U08956 U08957 U08958 U08959 U08960 U08962 U08963 U08964 U08965 U08966 U08967 U08968 U08969 U08970 U08971 U08972 U08973 U08974 U08975 U10019 U10020 U10021 U10022 U10023 U10024 U10025 U10026).

[Sala et al.(1995)] M. Sala, E. Pelletier, & S. Wain-Hobson. HIV-1 gp120 sequences from a doubly infected drug user. *AIDS Res Hum Retroviruses* **11**:653–655, 1995. OTE: (Medline: 96093902), (Genbank: U20670 U20671 U20672 U20673 U20674 U20675 U20676 U20677).

[Sala et al.(1994)] M. Sala, G. Zambruno, J. P. Vartanian, A. Marconi, U. Bertazzoni, & S. Wain-Hobson. Spatial discontinuities in human immunodeficiency virus type 1 quasispecies derived from epidermal Langerhans cells of a patient with AIDS and evidence for double infection. *J Virol* **68**:5280–5283, 1994. OTE: (Medline: 94309197).

[Salminen et al.(1996)] M. O. Salminen, B. Johansson, A. Sonnerborg, S. Ayehunie, D. Gotte, P. Leinikki, D. S. Burke, & F. E. McCutchan. Full-length sequence of an Ethiopian human immunodeficiency virus type i (HIV-1) isolate of genetic subtype C. *AIDS Res. Hum. Retroviruses* **12**:1329–1339, 1996.

[Salminen et al.(1995)] M. O. Salminen, C. Koch, E. Sanders-Buell, P. K. Ehrenberg, N. L. Michael, J. K. Carr, D. S. Burke, & F. E. McCutchan. Recovery of virtually full-length HIV-1 provirus of diverse subtypes from primary virus cultures using the polymerase chain reaction. *Virology* **213**:80–86, 1995. OTE: (Medline: 96036482), (Genbank: M17449 M19921 U26546 U26942).

[Sanchez-Pescador et al.(1985)] R. Sanchez-Pescador, M. D. Power, P. J. Barr, K. S. Steimer, M. M. Stempien, S. L. Brown-Shimer, W. W. Gee, A. Renard, A. Randolph, & J. A. L. et al. Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). *Science* **227**:484–92, 1985. OTE: (Medline: 85090453), (Genbank: K02007).

[Sauermann et al.(1990)] U. Sauermann, J. Schneider, J. Mous, U. Brunckhorst, I. Schedel, K. D. Jentsch, & G. Hunsmann. Molecular cloning and characterization of a German HIV-1 isolate. *AIDS Res Hum Retroviruses* **6**:813–23, 1990. OTE: (Medline: 90303973).

[Scarlatti et al.(1993)] G. Scarlatti, T. Leitner, E. Halapi, J. Wahlberg, P. Marchisio, M. A. Clerici-Schoeller, H. Wigzell, E. M. Fenyo, J. Albert, & M. U. et al. Comparison of variable region 3 sequences of human immunodeficiency virus type 1 from infected children with the RNA and DNA sequences of the virus populations of their mothers. *Proc Natl Acad Sci U S A* **90**:1721–5, 1993. OTE: (Medline: 93189569), (Genbank: L08277 L08278 L08279 L08280 L08281 L08282 L08283 L08284 L08285 L08286 L08287 L08288 L08289 L08290 L08291 L08292 L08293 L08294 L08295 L08296 L08297 L08298 L08299 L08300 L08301 L08302 L08303 L08304 L08305 L08306).

[Shapshak et al.(1995)] P. Shapshak, I. Nagano, K. Xin, W. Bradley, C. B. McCoy, N. C. Sun, R. V. Stewart, M. Yoshioka, C. Petito, & K. G. et al. HIV-1 heterogeneity and cytokines. Neuropathogenesis. *Adv Exp Med Biol* **373**:225–238, 1995. OTE: (Medline: 95397707).

[Shimizu et al.(1992)] H. Shimizu, F. Hasebe, H. Tsuchie, S. Morikawa, H. Ushijima, & T. Kitamura. Analysis of a human immunodeficiency virus type 1 isolate carrying a truncated transmembrane glycoprotein. *Virology* **189**:534–46, 1992. OTE: (Medline: 92351552), (Genbank: D01205 D01206 D01207 D01205 D01206 D01207 D01205 D01206 D01207).

[Shpaer et al.(1994)] E. G. Shpaer, E. L. Delwart, C. L. Kuiken, J. Goudsmit, M. H. Bachmann, & J. I. Mullins. Conserved V3 loop sequences and transmission of human immunodeficiency virus type 1. *AIDS Res Hum Retroviruses* **10**:1679–1684, 1994. OTE: (Medline: 95194734).

[Simmonds et al.(1990)] P. Simmonds, P. Balfe, C. A. Ludlam, J. O. Bishop, & A. J. Brown. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. *J Virol* **64**:5840–50, 1990. OTE: (Medline: 91056552), (Genbank: M36997).

[Siwka et al.(1994)] W. Siwka, A. Schwinn, K. Baczko, I. Pardowitz, F. Mhalu, J. Shao, A. Rethwilm, & M. ter V. vpu and env sequence variability of HIV-1 isolates from Tanzania. *AIDS Res Hum Retroviruses* **10**:1753–1754, 1994. OTE: (Medline: 95194744).

[Slobod et al.(1994)] K. S. Slobod, S. D. Rencher, A. Farmer, F. S. Smith, & J. L. Hurwitz. HIV type 1 envelope sequence diversity in inner city community. *AIDS Res Hum Retroviruses* **10**:873–875, 1994. OTE: (Medline: 95078006), (Genbank: U09140 U09141 U09142 U09143 U09144 U09145 U09146 U09147 U09148 U09149 U09150 U09151 U09158 U09159 U09160 U09161 U09162 U09163 U09164 U09165 U09166 U09167 U09168 U09169 U09170 U09171 U09172 U09173 U09174 U09175).

[Smith & Waterman(1992)] T. F. Smith & M. S. Waterman. The continuing case of the Florida dentist. *Science* **256**:1155–6, 1992. OTE: (Medline: 92271243).

[Sova et al.(1995)] P. Sova, M. van Ranst, P. Gupta, R. Balachandran, W. Chao, S. Itescu, G. McKinley, & D. J. Volsky. Conservation of an intact human immunodeficiency virus type 1 vif gene in vitro and in vivo. *J Virol* **69**:2557–2564, 1995. OTE: (Medline: 95191036).

[Spire et al.(1989)] B. Spire, J. Sire, V. Zachar, F. Rey, F. Barre-Sinoussi, F. Galibert, A. Hampe, & J. C. Chermann. Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the human immunodeficiency virus. *Gene* **81**:275–84, 1989. OTE: (Medline: 90034200).

[Srinivasan et al.(1987)] A. Srinivasan, R. Anand, D. York, P. Ranganathan, P. Feorino, G. Schochetman, J. Curran, V. S. Kalyanaraman, P. A. Luciw, & R. Sanchez-Pescador. Molecular characterization of human immunodeficiency virus from Zaire: nucleotide sequence analysis identifies conserved and variable domains in the envelope gene. *Gene* **52**:71–82, 1987. OTE: (Medline: 87248097), (Genbank: K03458 M16322).

[Srinivasan et al.(1989)] A. Srinivasan, D. York, D. B. Jr, R. Jannoun-Nasr, J. Getchell, J. McCormick, C. Y. Ou, G. Myers, T. Smith, & E. C. et al. Molecular characterization of HIV-1 isolated from a serum collected in 1976: nucleotide sequence comparison to recent isolates and generation of hybrid HIV. *AIDS Res Hum Retroviruses* **5**:121–9, 1989. OTE: (Medline: 89228766).

[Starcich et al.(1986)] B. R. Starcich, B. H. Hahn, G. M. Shaw, P. D. McNeely, S. Modrow, H. Wolf, E. S. Parks, W. P. Parks, S. F. Josephs, & R. C. G. et al. Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell* **45**:637–48, 1986. OTE: (Medline: 86218077), (Genbank: K03455 M12507 M12508 M17451).

[Stevenson et al.(1990)] M. Stevenson, S. Haggerty, C. Lamonica, A. M. Mann, C. Meier, & A. Wasiak. Cloning and characterization of human immunodeficiency virus type 1 variants diminished in the ability to induce syncytium-independent cytolysis. *J Virol* **64**:3792–803, 1990. OTE: (Medline: 90317877).

[Strunnikova et al.(1995)] N. Strunnikova, S. C. Ray, R. A. Livingston, E. Rubalcaba, & R. P. Viscidi. Convergent evolution within the V3 loop domain of human immunodeficiency virus type 1 in association with disease progression. *J Virol* **69**:7548–7558, 1995. OTE: (Medline: 96078998), (Genbank: U22682 U22683 U22684 U22685 U22686 U22687 U22688 U22689 U22690 U22691 U22692 U22693 U22694 U22695 U22696 U22697 U22698 U22699 U22700 U22701 U22702 U22703 U22704 U22705 U22706 U22707 U22708 U22709 U22710 U22711).

[Takeuchi et al.(1991)] Y. Takeuchi, M. Akutsu, K. Murayama, N. Shimizu, & H. Hoshino. Host range mutant of human immunodeficiency virus type 1: modification of cell tropism by a single point mutation at the neutralization epitope in the env gene. *J Virol* **65**:1710–8, 1991. OTE: (Medline: 91162716), (Genbank: M58037 M58038 M58039 M59192 M59193 M59888 M59889 M59890 M59891 M59892).

[Tan et al.(1993)] W. Tan, R. Fredriksson, A. Bjorndal, P. Balfe, & E. M. Fenyo. Cotransfection of HIV-1 molecular clones with restricted cell tropism may yield progeny virus with altered phenotype. *AIDS Res Hum Retroviruses* **9**:321–9, 1993. OTE: (Medline: 93290926).

[Tripathy et al.(1996)] S. P. Tripathy, B. Renjifo, W. Wang, M. F. McLane, R. Bollinger, J. Rodrigues, J. Osterman, S. Tripathy, & M. Essex. Envelope glycoprotein 120 sequences of primary HIV type 1 isolates from Pune and New Delhi, India. *AIDS Res. Hum. Retroviruses* **12**:1199–1202, 1996.

[Vanden Haesevelde et al.(1994)] M. Vanden Haesevelde, J. L. Decourt, L. D. R. J., B. Vanderborght, G. van der Groen, H. van Heuverswijn, & E. Saman. Genomic cloning and complete sequence analysis of a highly divergent African human immunodeficiency virus isolate. *J Virol* **68**:1586–96, 1994. OTE: (Medline: 94149849), (Genbank: L23119).

[von A. et al.(1995)] B. von A., B. von B., J. Eberle, B. Biryahwaho, R. G. Downing, & L. Gurtler. New crown motif of an HIV-1 V3 loop sequence from a Ugandan AIDS patient. *AIDS Res Hum Retroviruses* **11**:183–184, 1995. OTE: (Medline: 95251931), (Genbank: U15005 U15006 U15007).

[Wain-Hobson et al.(1985)] S. Wain-Hobson, P. Sonigo, O. Danos, S. Cole, & M. Alizon. Nucleotide sequence of the AIDS virus, LAV. *Cell* **40**:9–17, 1985. OTE: (Medline: 85099333), (Genbank: K02013).

[Wain-Hobson et al.(1991)] S. Wain-Hobson, J. P. Vartanian, M. Henry, N. Chenciner, R. Cheynier, S. Delassus, L. P. Martins, M. Sala, M. T. Nugeyre, & D. G. et al. LAV revisited: origins of the early HIV-1 isolates from Institut Pasteur. *Science* **252**:961–5, 1991. OTE: (Medline: 91240282), (Genbank: M64178 M64179 M64180 M64181 M64182 M64183 M64184 M64185 M64186 M64187 M64188 M64189 M64190 M64191 M64192 M64193 M64194 M64195 M64196 M64197 M64198 M64199 M64200 M64201 M64202 M64203 M64204 M64205 M64206 M64207).

[Wang et al.(1996a)] W. Wang, K. H. Mayer, M. Essex, & T. Lee. Sequential change of cysteine residues in hypervariable region 1 of gp120 in primary HIV-1 isolates of subtype B. *AIDS Res. Hum. Retroviruses* **12**:1195–1197, 1996a.

[Wang et al.(1996b)] W.-K. Wang, H. Brumblay, M. Essex, & T.-H. Lee. The ratio of non-synonymous nucleotide substitutions to synonymous nucleotide substitutions per site of divergent gp120 sequences: Intrasubject and intersubject comparisons. *Unpublished* 1996b.

[Wang et al.(1995)] W. K. Wang, M. Essex, & T. H. Lee. Uncommon gp120 cysteine residues found in primary HIV-1 isolates. *AIDS Res Hum Retroviruses* **11**:185–188, 1995. OTE: (Medline: 95251932), (Genbank: U16324 U16325 U16326 U16327 U16328 U16329 U16330 U16331 U16332 U16333 U16334 U16335).

[Weniger et al.(1994)] B. G. Weniger, Y. Takebe, C. Y. Ou, & S. Yamazaki. The molecular epidemiology of HIV in Asia. *AIDS* **8**:S13–28, 1994. OTE: (Medline: 95160969), (Genbank: L07442 L07443 L07444 L07445 L07446 L07447 L07448 L07449 L07450 L07451 L07452 L07453 L07454 L07455 L07456 L07457 L07458 L07459 L07460 L07461 L07462 L07463 L07464 L19238 L19239 L32084 L32085 L32086 L32087 L32088).

[Westervelt et al.(1991)] P. Westervelt, H. E. Gendelman, & L. Ratner. Identification of a determinant within the human immunodeficiency virus 1 surface envelope glycoprotein critical for productive infection of primary monocytes. *Proc Natl Acad Sci U S A* **88**:3097–101, 1991. OTE: (Medline: 91195299), (Genbank: M60157 M60158 M60159 M60160 M60161 M60162 M60163 M60164 M60165 M60472).

[Willey et al.(1986)] R. L. Willey, R. A. Rutledge, S. Dias, T. Folks, T. Theodore, C. E. Buckler, & M. A. Martin. Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome retrovirus. *Proc Natl Acad Sci U S A* **83**:5038–42, 1986. OTE: (Medline: 86259728), (Genbank: K03346 K03347 K03455).

[Wolfs et al.(1992)] T. F. Wolfs, G. Zwart, M. Bakker, & J. Goudsmit. HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. *Virology* **189**:103–10, 1992. OTE: (Medline: 92295544), (Genbank: M91819 M91820 M91821 M91822 M91823 M91824 M91825 M91826 M91819 M91820 M91821 M91822 M91823 M91824 M91825 M91826).

[Wolfs et al.(1991)] T. F. Wolfs, G. Zwart, M. Bakker, M. Valk, C. L. Kuiken, & J. Goudsmit. Naturally occurring mutations within HIV-1 V3 genomic RNA lead to antigenic variation dependent on a single amino acid substitution. *Virology* **185**:195–205, 1991. OTE: (Medline: 92024075), (Genbank: M74591 M74592 M74593 M74594 M74595 M74596 M74597 M74598 M74599 M74600 M74601 M74602 M74603 M74604 M74605 M74606 M74607 M74608 M74609 M74610 M74611 M74612 M74613 M74614 M74615 M74616 M74617 M74618 M74619 M74620).

[Wolynsky et al.(1996)] S. M. Wolynsky, B. T. M. Korber, A. U. Neumann, M. Daniels, K. J. Kunstman, A. J. Whetsell, Y. Cao, D. D. Ho, J. T. Safrit, & R. A. Koup. Adaptive evolution of human immunodeficiency virus type 1 during the natural course of infection. *Science* **272**:537–542, 1996.

[Wrin et al.(1995)] T. Wrin, T. P. Loh, J. C. Vennari, H. Schuitemaker, & J. H. Nunberg. Adaptation to persistent growth in the H9 cell line renders a primary isolate of human immunodeficiency virus type 1 sensitive to neutralization by vaccine sera. *J Virol* **69**:39–48, 1995. OTE: (Medline: 95074890), (Genbank: U15030 U15032).

[Xin et al.(1995)] K. Q. Xin, P. Shapshak, S. Kawamoto, I. Nagano, C. B. McCoy, & K. Okuda. Highly divergent env sequences of HIV-1 B subtype with two novel V3 loop motifs detected in an AIDS patient in Miami, Florida. *AIDS Res. Hum. Retroviruses* **11**:1139–41, 1995. OTE: (Medline: 96089223).

[York-Higgins et al.(1990)] D. York-Higgins, C. Cheng-Mayer, D. Bauer, J. A. Levy, & D. Dina. Human immunodeficiency virus type 1 cellular host range, replication, and cytopathicity are linked to the envelope region of the viral genome. *J Virol* **64**:4016–20, 1990. OTE: (Medline: 90317906).

[Yourno et al.(1988)] J. Yourno, S. F. Josephs, M. Reitz, D. Zagury, F. Wong-Staal, & R. C. Gallo. Nucleotide sequence analysis of the env gene of a new Zairian isolate of HIV-1. *AIDS Res Hum Retroviruses* **4**:165–73, 1988. OTE: (Medline: 88281278), (Genbank: J03653).

[Yu et al.(1995)] X. F. Yu, Z. Wang, C. Beyrer, D. D. Celentano, C. Khamboonruang, E. Allen, & K. Nelson. Phenotypic and genotypic characteristics of human immunodeficiency virus type 1 from patients with AIDS in northern Thailand. *J Virol* **69**:4649–4655, 1995. OTE: (Medline: 95333239), (Genbank: L22943 L22957 M74978 M74979 M74980 M74981 M74982 M74983 M74984 M74985 M74986 M74987 M74988 M74989 M74990 M74991 M74992 M74993 M74994 M74995 M74996 M74997 M74998 M74999 M75000 M75001 M75002 M75003 M75004 M75005).

[Zachar et al.(1996a)] V. Zachar, A. S. Goustin, V. Zacharova, H. Hager, U. Koppelhus, D. D. Womble, X. Liu, C. Bambra, A. Nyongo, & P. Ebbesen. Genetic polymorphism of envelope V3 region of HIV type 1 subtypes A, C, and D from Nairobi, Kenya. *AIDS Res. Hum. Retroviruses* **12**:75–78, 1996a.

[Zachar et al.(1996b)] V. Zachar, V. Mayer, V. Zacharova, H. Schmidtmayerova, J. Kasanicka, M. Mokras, X. Liu, A. S. Goustin, & P. Ebbesen. Spread of HIV type 1 in Slovakia remains limited and is restricted to subtype B. *AIDS Res. Hum. Retroviruses* **12**:1069–1071, 1996b.

[Zhang et al.(1993)] L. Q. Zhang, P. MacKenzie, A. Cleland, E. C. Holmes, A. J. Brown, & P. Simmonds. Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. *J Virol* **67**:3345–56, 1993. OTE: (Medline: 93267786), (Genbank: L13488 L13489 L13490 L13491 L13492 L13493 L13494 L13495 L13496 L13497 L13498 L13499 L13500 L13501 L13502 L13503 L13504 L13505 L13506 L13507 L13508 L13509 L13510 L13511 L13512 L13513 L13514 L13515 L13516 L13517).

[Zhu & Ho(1995)] T. Zhu & D. D. Ho. Was hiv present in 1959? *Nature* **374**:503–4, 1995. OTE: (Medline: 95214770).

[Zhu et al.(1993)] T. Zhu, H. Mo, N. Wang, D. S. Nam, Y. Cao, R. A. Koup, & D. D. Ho. Genotypic and phenotypic characterization of HIV-1 patients with primary infection. *Science* **261**:1179–81, 1993. OTE: (Medline: 93361976), (Genbank: L21224 L21225 L21226 L21227 L21228 L21229 L21230 L21231 L21232 L21233 L21234 L21235 L21236 L21237 L21238 L21239 L21240 L21241 L21242 L21243 L21244 L21245 L21246 L21247 L21248 L21249 L21250 L21251 L21252 L21253).

[Zhu et al.(1995)] T. Zhu, N. Wang, A. Carr, S. Wolinsky, & D. D. Ho. Evidence for coinfection by multiple strains of human immunodeficiency virus type 1 subtype B in an acute seroconvertor. *J Virol* **69**:1324–1327, 1995. OTE: (Medline: 95115099), (Genbank: U16372 U16373 U16374 U16375 U16376 U16377 U16378 U16379 U16380 U16381 U16382 U16383 U16384 U16385 U16386 U16387 U16388).

[Zwart et al.(1994a)] G. Zwart, N. K. Back, C. Ramautarsing, M. Valk, L. van der Hoek, & J. Goudsmit. Frequent and early HIV-1MN neutralizing capacity in sera from Dutch HIV- 1 seroconverters is related to antibody reactivity to peptides from the gp120 V3 domain. *AIDS Res Hum Retroviruses* **10**:245–251, 1994a. OTE: (Medline: 94289062).

[Zwart et al.(1994b)] G. Zwart, L. van der Hoek, M. Valk, M. T. Cornelissen, E. Baan, J. Dekker, M. Koot, C. L. Kuiken, & J. Goudsmit. Antibody responses to HIV-1 envelope and gag epitopes in HIV-1 seroconverters with rapid versus slow disease progression. *Virology* **201**:285–93, 1994b. OTE: (Medline: 94240812), (Genbank: U05786 U05787 U05788 U05789 U05790 U05791 U05792 U05793 U05794 U05795 U05796 U05797 U05798 U05799 U05800 U05801 U05802 U05803 U05804 U05805 U05806 U05807 U05808 Z29279 Z29287 Z29289 Z29292).

[Zwart et al.(1993)] G. Zwart, T. F. Wolfs, R. Bookelman, S. Hartman, M. Bakker, C. A. Boucher, C. Kuiken, & J. Goudsmit. Greater diversity of the HIV-1 V3 neutralization domain in Tanzania compared with The Netherlands: serological and genetic analysis. *AIDS* **7**:467–74, 1993. OTE: (Medline: 93283054).

[Zwart et al.(1992)] G. Zwart, T. F. Wolfs, M. Valk, L. Van der Hoek, C. L. Kuiken, & J. Goudsmit. Characterization of the specificity of the human antibody response to the V3 neutralization domain of HIV-1. *AIDS Res Hum Retroviruses* **8**:1897–908, 1992. OTE: (Medline: 93143999).

Structural Studies Involving Different HIV-1 V3 Loops

Paolo Catasti and Goutam Gupta

MS K710, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Introduction

Studies on the feasibility of a subunit vaccine to protect against human immunodeficiency virus (HIV) infection have principally focused on the third variable (V3) loop of the envelope surface protein. One of the neutralizing determinants of HIV-1 is located inside the V3 loop. However, progress toward a vaccine based on neutralizing determinants has been impeded by the amino acid sequence variability in the V3 loop of different HIV isolates. The elusive nature of the V3 loop structure prompted us to carry out a systematic study on different isolates in an attempt to identify a common structural motif in the V3 loop regardless of the amino acid sequence variability. We have performed 2D NMR structural studies on three different V3 loop peptides: MN, Haiti (Haiti 6004; L07201), and RF (Catasti *et al.*, 1995 & 1996). The three V3 loops were all 35 residues long and S-S bridged at the terminals. The NMR studies were carried out first in water, then in a 70%/30% mixture of water/trifluoroethanol 1 (TFE). TFE is a solvent widely used in NMR on peptides, for its property to unmask helical propensities of hydorphobic residues.

Figure 1 shows that similar secondary structures are observed for the three different V3 loops: a GPG(K/R) crest in the center of the neutralizing determinant, two extended regions flanking the central crest, and a helical region in the C-terminal domain observed only in the water/TFE mixture. The RF V3 peptide did not dissolve in the water/TFE mixture, therefore we could run the experiments only in an aqueous solution. Structural prediction studies revealed that the variability in sequence and structure of the V3 loop is confined to the N and C-terminal side of the conserved GPG crest. Figure 2 is a summary of the NMR secondary structural assignments (Catasti *et al.*, 1995 & 1996), and the results of several secondary prediction algorithms. With the exception of the PSA method, most of the algorithms fail to identify the alpha helix in the C-terminal portion of the V3 loops.

References

Catasti, P., Bradbury, E.M., and Gupta G. (1996) Structure and polymorphism of HIV-1 third variable loops. *J. Biol. Chem.*, **271**(14):8236–8242.

Catasti, P., Fontenot, J.D., Bradbury, E.M., and Gupta G. (1995) Local and global structural properties of the HIV-MN V3 loop. *J. Biol. Chem.*, **270**(5):2224–2232.

Deleage, G., and Roux, B. (1987) An algorithm for protein secondary strcuture prediction based on class prediction. *Prot. Eng.*, **1**:289–294.

Geourjon, C., and Deleage, G. (1994) SOPM : a self optimised prediction method for protein secondary structure prediction. *Prot. Eng.*, **7**:157–164.

Geourjon, C., and Deleage, G. (1995) SOPMA : Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *CABIOS*, **11**:681–684

Gibrat, J.F., Garnier, J., and Robson, B. (1987) Further developments of protein secondary structure prediction using information theory. New parameters and consideration of residue pairs. *J. Mol. Biol.*, **198**(3):425–443.

King, R.D., and Sternberg, M.J.E. (1996) Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Prot. Science*, in press.

Kneller, D.G., Cohen, F.E., and Langridge, R. (1990) Improvements in Protein Secondary Structure Prediction by an Enhanced Neural Network. *J. Mol. Biol.* **214**:171–182.

Levin, J.M., Robson, B., and Garnier, J. (1986) An algorithm for secondary structure determination in proteins based on sequence similarity. *Febs Lett.*, **205**(2):303–308.

McClelland, J.L., and Rumelhart, D.E. (1988) *Explorations in Parallel Distributed Processing* vol 3. pp 318–362. MIT Press, Cambridge MA.

Rost, B. (1996). PHD: predicting one-dimensional protein structure by profile based neural networks. *Meth. in Enzym.*, **266**:525–539.

Stultz, C.M., White, J.V., and Smith, T.F. (1993) Structural Analysis Based on State-space Modeling. *Prot. Science*, **2**:305–314.

White, J.V., Stultz, C.M., and Smith, T.F. (1994) Protein Classification by Stochastic Modeling and Optimal Filtering of Amino-Acid Sequences. *Math. Biosc.*, **119**:35–75.

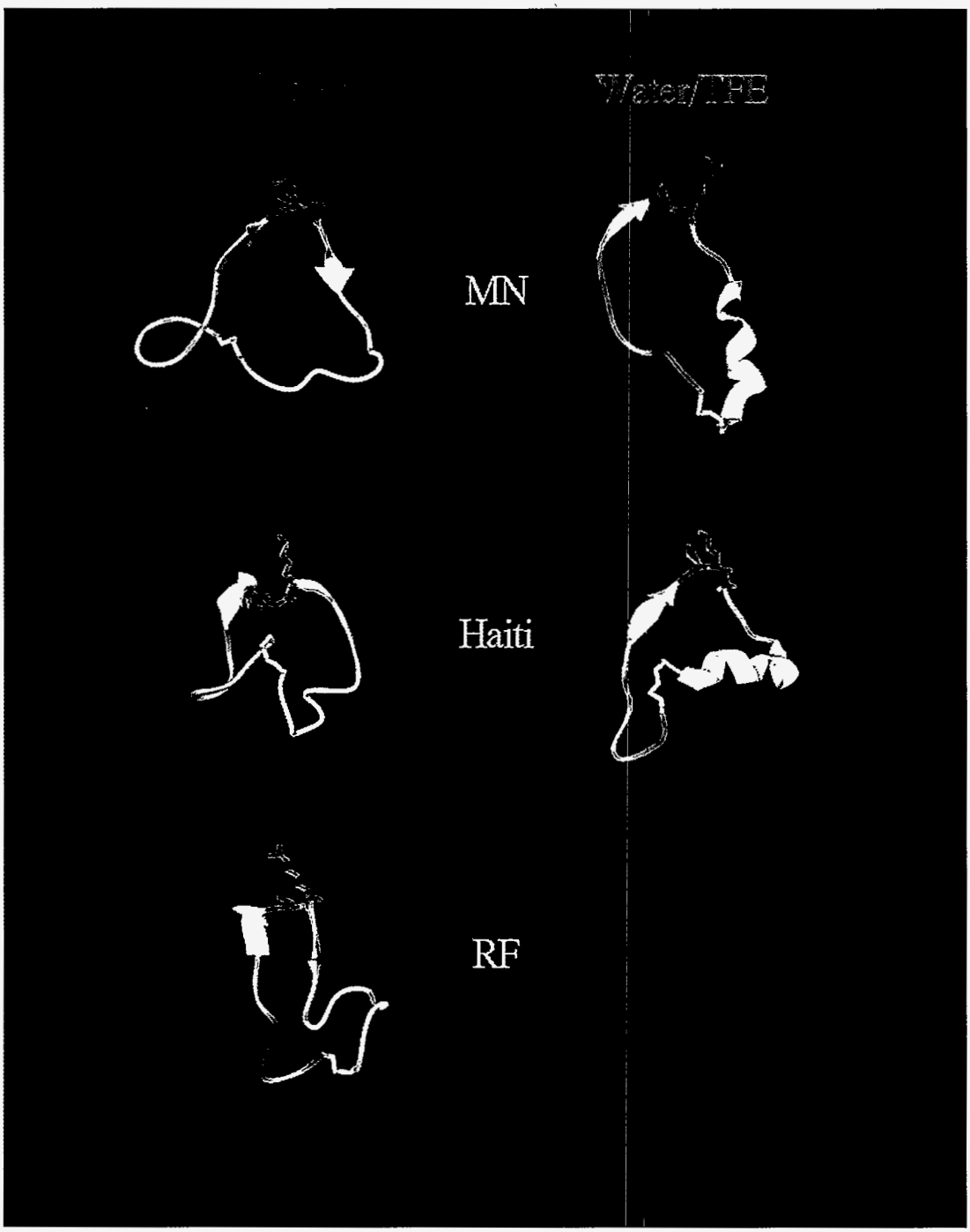


Figure 1. Ribbon diagram showing the average folding patterns of the structures of the MN, Haiti and RF V3 loops in water and in a mixture of 70%/30% water/TFE. In each case the average is done over 70 sampled low energy structures. Note that, in each case, the neutralizing epitope containing the central GPG(R/K) sequence forms a protruding loop even though the local structure and presentation of the loop in the different cases are noticeably different. Structures that satisfy the NMR constraints of the V3 loops in water show a higher degree of flexibility than those in agreement with the NMR data in the mixed water/ TFE solvent. This is due to the formation of the alpha helix in the mixed solvent. Color code is as follows: GPG(R/K) crest is red, extended regions are green, disulfide bridges are yellow and the alpha helical region is cyan.

```
                    1       10        20        30
                    |       |         |         |
V3MN                CTRPNYNKRKRIHIGPGRAFYTTKNIIGTIRQAHC
NMR (water)         .EECTTTTCCEEEETTTTEEEECCCCTTTTTTCC.
NMR (TFE)           .EECTTTTEEEECCTTTTCCCCCHHHHHHHHHHC.
PSA method          CCCCCCCCCCCEEEECCTCCCCCCCHHHHHHHHHHH
Gibrat method       ECCCCCCCCEEEEEECCCCEEEEEEEHEEEEEECCC
Levin method        .TSSSSCSSSCEEECTTCCEEEEECCCCHHHHSSC
DPM method          CCCTTCTCHHHECCTCCCHHCEECCEEEEEHHHCC
SOPMA predict       ECCTTHCTEEEEEEECCCCCCTTTCCCCTTCCCCE
DSC method          CCCCCCCCCCEEEEECCCCCEEECCCCCCCCCCCCC
PHD method          CCCCCCCCCEEEE.CCC.EE..CC........CCC
nnpredict method    CCCCCCCCCCCCCCCCCCEEEECCCCEEECHHCCCC
                    1       10        20        30
                    |       |         |         |
V3Haiti             CTRPNDNTRKSIPMGPGKAFYATGDIIGNIRQAHC
NMR (water)         .CECTTTTEEECCCTTTTEECCCCCCTTTTTTTC.
NMR (TFE)           .EECTTTTCCEEEETTTTCCTTTTCHHHHHHHHH.
PSA method          CCCCCCCCCCCCCCCCCCCCCCCCCHHCCHHHHHC
Gibrat method       ECCCCCCCCCCCCCCCCCCEEEEEHHEEEEEECCE
Levin method        TCCCCCCSSCCCCCCSTCCECCTSCCSCSCCHCCC
DPM method          CCCCTCTCCCCCCCTCCCHHCHECCEECCEHHHCC
SOPMA predict       ECCCCTTCCCCETTTCHHHHHETHHHECCCTTCCC
DSC method          CCCCCCCCCCCCCCCCCCEEECCCCCCCCCCCCCC
PHD method          CCCCCCCC.....CCCC.EE.CCC.........C
nnpredict method    CCCCCCCCCCCCCCCCCCCEECCCCCEECHHCCCC
                    1       10        20        30
                    |       |         |         |
V3RF                CTRPNNNTRKSITKGPGRVIYATGQIIGDIRKAHC
NMR (water)         .CTTTTEECEECCCTTTTEECCCTTTTCCTTTTC.
PSA method          CCCCCCCCCCCCCCCCCCEEEECCCHHHHHHHHHH
Gibrat method       ECCCCCCCCCEEEECCCCEEEEEECEECEEEEEHCC
Levin method        CCCCCCCSHS..CCCTTCEEEECCCCCCSCCCCC.
DPM method          CCTTTCTCCCCECCTCCCEEEEECCEECCEHHHCC
SOPMA predict       ECCCCCCCEEEECCCCCCCCEEETTCEEEHHHHHHH
DSC method          CCCCCCCCCCEEECCCCCEEEEECCCCCCCCCCCCC
PHD method          CCCCCCCC..E..CCCC.EEE.CC.EE.......C
nnpredict method    CCCCCCCCCCCCCCCCCCEEEECCCCEEECHHCCCC
```

Figure 2. Comparison of secondary structure assignments of the NMR determined structures and secondary structure prediction for the three V3 loops, MN, Haiti and RF. The different prediction algorithms are indicated on the left. Some of these methods are discussed by Myers and Farmer in Part III of this compendium.

Meaning of Symbols

H	alpha helix	T	turn
C	random coil/loop	S	bend
E	strand	.	unassigned

Key to Prediction Algorithms

PSA	Stultz *et al.*	Gibrat	Gibrat *et al.*
Levin	Levin *et al.*	DPM	Deleage *et al.*
SOPM	Geourjon *et al.*	DSC	King *et al.*
PHD	Rost *et al.*	nnpredict	McClelland *et al.*

Mutations in Retroviral Genes Associated with Drug Resistance

John W. Mellors[1] , Raymond F. Schinazi[2] , and Brendan A. Larder[3]

[1] *University of Pittsburgh Medical Center, Monteflore University Hospital, Infectious Diseases Division, 200 Lothrop Street, Pittsburgh, PA 15213-2582.*
[2] *Emory University/VAMC. 1670 Clairmont Rd., Decatur, GA 30033.*
[3] *Glaxo Wellcome Medicines Research Centre, Gunnels Wood Rd., Stevenage, Herts, SG1 2NY, UK.*

Introduction

The emergence of drug-resistant variants of HIV continues to be of prime interest in the fields of HIV disease pathogenesis and antiretroviral chemotherapy. Drug resistance is the inevitable consequence of incomplete suppression of HIV replication. The rapid replication rate of HIV and its inherent genetic variation lead to the generation of a seemingly limitless number of viral variants that exhibit drug resistance. The growing number of drug resistance mutations listed in these revised tables stands as a testimony to the genetic flexibility of HIV. When the first resistance table was published in 1994 (*International Antiviral News* 2(5):72–75) only 42 different mutations were listed. This new update lists 143 mutations, a 240 per cent increase over a 2-year period. The revised tables include, for the first time, sections on HIV binding/fusion inhibitor resistance, and multidrug resistance. Although the tables are quite comprehensive, the reader should be reminded that the mutations described are predominately found in clade B virus and not in other HIV genotypes.

These tables are presented here in a modified format courtesy of the International Medical Press, which previously published them in *International Antiviral News* August, 1996 4(6):95–107.

Acknowledgments

The authors would like to gratefully acknowledge their colleagues for assistance in assembling this table, and in particular the following: E Anton, J Balzarini, CAB Boucher, RW Buckheit, JM Cherrington, J Condra, R D'Aquila, MP de Bethune, K De Vreese, E De Clercq, LM Demeter, R Deziel, JA Este, J Eberle, E Emini, JW Erickson, A Fridland, S Gulnik, D Havlir, JP Kleim, D Lamarre, P-F Lin, TC Merigan, M Moermans, A Molla, T North, AK Patick, W Shao, S Sharma, R Swanstrom, A-M Vandamme, K Van Rompay, MA Wainberg, G Werner, SS Yang, SD Young, S Yuasa, and H Zhang. The authors wish to thank A Juodawlkis for her excellent assistance with the table. This work was supported in part by the Department of Veterans Affairs and the Georgia VA Research Center for AIDS and HIV Infections.

Mutations in HIV-1 RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
M 41 L	ATG to TTG/CTG	Nucleoside RTI	AZT	?	Y	4		M41L/T215Y: 60-70-fold; M41L/D67N/ K70R/T215Y: 180-fold.	(1,2,3)
K 65 R	AAA to AGA	Nucleoside RTI	1592U89	Y	N	3		K65R/L74V and/or Y115F with M184V: 10 fold; K65R/M184V: 8-fold	(4,5)
K 65 R	AAA to AGA	Nucleoside RTI	ddC	Y	Y	4-10			(6,7)
K 65 R	AAA to AGA	Nucleoside RTI	ddI	Y	Y	4-10	ddI; ddC; PMEA; 3TC	Infrequently observed in patients receiving ddI or ddC	(6,7)
K 65 R	AAA to AGA	Nucleoside RTI	DXG	Y	?	8	other dioxolane derivatives	K65R reverses AZT resistance in D67N/K70R/T215Y/K219Q background	(8,9)
K 65 R	AAA to AGA	Nucleoside RTI	PMEA	Y	?	10-25			(10,11)
D 67 N	GAC to AAC	Nucleoside RTI	AZT	Y	Y			D67N/K70R/T215Y/K219Q: 120-fold; M41L/D67N/K70R/T215Y: 180-fold.	(1,2,3)
T 69 D	ACT to GAT	Nucleoside RTI	ddC	N	Y	5			(12)
K 70 E	AAA to GAA	Nucleoside RTI	PMEA	Y	?	9	3TC; PFA: 2-fold hypersusceptibility		(13)
K 70 R	AAA to AGA	Nucleoside RTI	AZT	Y	Y			D67N/K70R/T215Y/K219Q: 120-fold	(1,2,3)
L 74 I	TTA to ATA	HIV-1 Specific RTI	HBY 097	Y	?				
L 74 V	TTA to GTA	HIV-1 Specific RTI	HBY 097	Y	?				
L 74 V	TTA to GTA	Nucleoside RTI	1592U89	Y	N	4		K65R/L74V and/or Y115F with M184V: 10 fold; L74V/M184V: 9-fold resistance; L74V/Y115F/M184V: 11-fold	(4,5)
L 74 V	TTA to GTA	Nucleoside RTI	ddI	N	Y	5-10	ddC	Can reverse effect of T215Y AZT resistance mutation	(14)
L 74 V	TTA to GTA	Nucleoside RTI	DXG	Y	?	4			(9)
V 75 I	GTA to TTA	HIV-1 Specific RTI	HBY 097	Y	?			Compensates for negative effect of the G190E mutation on RT activity	(15)
V 75 L	GTA to ATA	HIV-1 Specific RTI	HBY 097	Y	?				

Mutations in HIV-1 RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
V 75 T	GTA to ACA	Nucleoside RTI	d4T	Y	Y	7	ddI; ddC; d4C; (-)-FTC	Observed with d4T selection in vitro, and rarely in patients receiving d4T	(16,17)
W 88 G	TGG to GGG	Pyrophosphate Analogue RTI	Foscarnet (PFA)	Y	Y	5	Hypersusceptibility to AZT	Observed after selection with AZT and PFA; suppresses effects of AZT mutations while maintaining 3.5- to 4.7-fold PFA resistance	(18,19)
W 88 S	TGG to TCG	Pyrophosphate Analogue RTI	Foscarnet (PFA)	N	Y	2-4	Wild-type susceptibility to AZT	Partially suppresses effects of AZT mutations, but resistance to PFA is lost	(18,20)
E 89 G	GAA to GGA	Pyrophosphate Analogue RTI	Foscarnet (PFA)	Y	N	14		Isolated by screening RT clones for ddGTP resistance	(21)
E 89 K	GAA to GGA	Pyrophosphate Analogue RTI	Foscarnet (PFA)	Y	N	> 16		Suppresses effects of AZT resistance mutations	(19)
L 92 I	TTA to ATA	Pyrophosphate Analogue RTI	Foscarnet (PFA)	Y	N	8		Suppresses effects of AZT resistance mutations	(19)
A 98 G	GCA to GGA	HIV-1 Specific RTI	L-697,661	N	Y	8			(22)
A 98 G	GCA to GGA	HIV-1 Specific RTI	Nevirapine	N	Y				(23)
L 100 I	TTA to ATA	HIV-1 Specific RTI	BHAP U-88204E	Y	?				(24,25)
L 100 I	TTA to ATA	HIV-1 Specific RTI	DMP 266 (L-743,726)	Y	?	8-11		Combinations of mutations needed for high-level resistance; L100I/V108I: 1,000-fold; L100I/V179D/Y181C: 1,000-fold	(26,27,28)
L 100 I	TTA to ATA	HIV-1 Specific RTI	L-697,661	Y	N	2			(22)
L 100 I	TTA to ATA	HIV-1 Specific RTI	Nevirapine	N	Y				(29)
L 100 I	TTA to ATA	HIV-1 Specific RTI	TIBO R82150	Y	?	> 100		Can reverse effects of AZT resistance mutations	(30,31,32)

Mutations in HIV-1 RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
L 100 I	TTA to ATA	HIV-1 Specific RTI	TIBO R82913	Y	?			Found in combination with E138K	(31,33)
L 100 I	TTA to ATA	HIV-1 Specific RTI	UC-68 (638532)	Y	?	70			(34)
L 100 I	TTA to ATA	HIV-1 Specific RTI	UC-70 (638534)	Y	?	758			(35)
L 100 I	TTA to ATA	HIV-1 Specific RTI	UC-781	Y	?	20		Activity of UC-781 versus L100I, K103N, V106A, E138K, Y181C and Y188L reduced by 2-, 7-, 1.5-, 1.5-, 5- and 150-fold, respectively, compared to wild type	(36,37)
L 100 I	TTA to ATA	HIV-1 Specific RTI	UC-84 (615985)	Y	?	> 40, > 33			(35,38)
K 101 E	AAA to GAA	HIV-1 Specific RTI	8-Chloro-TIBO R091767	?	Y				(39)
K 101 E	AAA to GAA	HIV-1 Specific RTI	BHAP U-87201E (atevirdine)	N	Y			K101E, Y188H, E233Y and K238T observed with U-87201E/AZT combination therapy	(40,41)
K 101 E	AAA to GAA	HIV-1 Specific RTI	DMP 266 (L-743,726)	Y	?	1,000			(26,27)
K 101 E	AAA to GAA	HIV-1 Specific RTI	L-697,661	N	Y	8			(22)
K 101 E	AAA to GAA	HIV-1 Specific RTI	UC-10 (645129)	Y	?			K101E/Y181C: 200-fold	(35)
K 101 E	AAA to GAA	HIV-1 Specific RTI	UC-38 (629243)	Y	?			K101E/G190E: > 100-fold; cross resistance to: TSAO-m^3 T, Nevirapine, TIBO R82913, BHAP U88204; susceptible to L697,661	(34,42)
K 101 E	AAA to GAA	HIV-1 Specific RTI	UC-57 (647014)	Y	?			K101E/Y181C: 58-fold	(35)
K 101 I	AAA to ATA	HIV-1 Specific RTI	UC-16	Y	?	10		K101I/G141E: 10-fold	(34)
K 101 Q	AAA to CAA	HIV-1 Specific RTI	Trovirdine	Y	?			Found in combination with V108I	(43,44)
K 103 N	AAA to AAC	HIV-1 Specific RTI	8-Chloro-TIBO R091767	?	Y				(39)
K 103 N	AAA to AAC	HIV-1 Specific RTI	BHAP U-87201E (atevirdine)	N	Y			K103N and Y181C observed with monotherapy	(40)

Mutations in HIV-1 RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
K 103 N	AAA to AAC	HIV-1 Specific RTI	BHAP U-90152 (delavirdine)	?	Y			K103N/Y181C seen with monotherapy and combination therapy	(45)
K 103 N	AAA to AAC	HIV-1 Specific RTI	DMP 266 (L-743,726)	Y	?	67			
K 103 N	AAA to AAC	HIV-1 Specific RTI	L-697,593	Y	?	20		K103N/Y181C: > 1,000-fold	(46)
K 103 N	AAA to AAC	HIV-1 Specific RTI	L-697,661	Y	Y	8		K103N and Y181C most common with monotherapy	(22,47)
K 103 N	AAA to AAC	HIV-1 Specific RTI	Loviride (R89439, α-APA)	Y	Y				(48)
K 103 N	AAA to AAC	HIV-1 Specific RTI	MKC-442 (I-EBU)	Y	?				(49)
K 103 N	AAA to AAC	HIV-1 Specific RTI	Nevirapine	N	Y				(29)
K 103 N	AAA to AAC	HIV-1 Specific RTI	TIBO R82913	Y	?	> 100		K103N/Y181C: > 1,000-fold	(24)
K 103 N	AAA to AAC	HIV-1 Specific RTI	UC-10 (645129)	Y	?	5			(34)
K 103 N	AAA to AAC	HIV-1 Specific RTI	UC-81 (615727)	Y	?	40			
K 103 Q	AAA to CAA	HIV-1 Specific RTI	L-697,661	N	Y	8			(47)
K 103 R	AAA to AGA	HIV-1 Specific RTI	Trovirdine	Y	?		Nevirapine; 9-chloro-TIBO	K103R/V179D: 500-fold; Found in combination with V179D or Y181C	(43,44)
K 103 T	AAA to ACA	HIV-1 Specific RTI	BHAP U-90152 (delavirdine)	?	Y				(45)
K 103 T	AAA to ACA	HIV-1 Specific RTI	UC-42	Y	N	100			(34)
V 106 A	GTA to GCA	HIV-1 Specific RTI	BHAP U-88204E	Y	?				(25)
V 106 A	GTA to GCA	HIV-1 Specific RTI	E-EBU-dM	Y	?				(50)
V 106 A	GTA to GCA	HIV-1 Specific RTI	Nevirapine	Y	Y	100		No effect on AZT resistance	(23,24,29,33)
V 106 A	GTA to GCA	HIV-1 Specific RTI	TIBO R82913	Y	?	100		No effect on AZT resistance	(33)
V 106 A	GTA to GCA	HIV-1 Specific RTI	UC-69 (646989)	Y	?			V106A/V181C: 166-fold	(35)

Mutations in HIV-1 RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
V 106 A	GTA to GCA	HIV-1 Specific RTI	UC-82	Y	?	13		Activity of UC-82 versus L100I, K103N, V106A, E138K, Y181C and Y188L reduced by 2-, 6-, 1.5-, 2-, 4- and 200-fold, respectively, compared to wild type	(36,37)
V 106 I	GTA to ATA	HIV-1 Specific RTI	HBY 097					Appears under lowered drug concentration selection	(51)
V 108 I	GTA to ATA	HIV-1 Specific RTI	DMP 266 (L-743,726)	Y	?			L100I/V108I: 1,000-fold	(26,27)
V 108 I	GTA to GCA	HIV-1 Specific RTI	L-697,661	Y	Y	4			(22)
V 108 I	GTA to ATA	HIV-1 Specific RTI	Loviride (R89439, α-APA)	Y	?				(48)
V 108 I	GTA to GCA	HIV-1 Specific RTI	MKC-442 (I-EBU)	Y	?				(49)
V 108 I	GTA to ATA	HIV-1 Specific RTI	Nevirapine	N	Y				(29)
V 108 I	GTT to GAT	HIV-1 Specific RTI	TIBO R82913	N	Y	>100	R82150 (>100)		(52)
V 108 I	GTA to ATA	HIV-1 Specific RTI	Trovirdine	Y	?			Found in combination with K101Q	(43,44)
Y 115 F	TAT to TTT	Nucleoside RTI	1592U89	Y	N	2		K65R/L74V and/or Y115F with M184V: 10 fold; L74V/Y115F/M184V: 11-fold	(4,5)
E 138 K	GAG to AAG	HIV-1 Specific RTI	TSAO	Y	?	> 100		E138A (GAG to GCG) in TSAO-naive patients confers TSAO resistance	(53,54,55)
E 138 K	GAG to AAG	HIV-1 Specific RTI	MKC-442 (I-EBU)	Y	N		TSAOs	Obtained in the concomitant presence of low 3TC concentrations	(56)
E 138 K	GAG to AAG	HIV-1 Specific RTI	TIBO R82913	Y	?		TSAOs	Found in combination with L100I	(31)
E 138 K	GAG to AAG	HIV-1 Specific RTI	UC-82	Y	?	5	TSAOs	Activity of UC-82 versus L100I, K103N, V106A, E138K, Y181C and Y188L reduced by 2-, 6-, 1.5-, 2-, 4- and 200-fold, respectively, compared to wild type	(36,37)

Mutations in HIV-1 RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
E 138 K	GAG to AAG	HIV-1 Specific RTI	UC-84 (615985)	Y	?	> 100	TSAOs		(34,57)
T 139 I	ACA to ATA	HIV-1 Specific RTI	Calanolide A	Y	?	> 70	Not other NNRTIs		(38)
G 141 E	GGG to GAG	HIV-1 Specific RTI	UC-16	Y	?			K101I/G141E: 10-fold	(34)
S 156 A	TCA to GCA	Pyrophosphate Analogue RTI	Foscarnet (PFA)	Y	N	4.5			(19)
Q 161 L	CAA to CTA	Pyrophosphate Analogue RTI	Foscarnet (PFA)	Y	Y	5		Q161L/H208Y: 9-fold; Q161L/H208Y reverses effects of AZT mutations D67N/K70R/T215Y/K219Q	(18)
V 179 D	GTT to GAT	HIV-1 Specific RTI	DMP 266 (L-743,726)	Y	?			L100I/V179D/Y181C: 1,000-fold	(26,27)
V 179 D	GTT to GAT	HIV-1 Specific RTI	L-697,661	N	Y	4			(22)
V 179 D	GTT to GAT	HIV-1 Specific RTI	TIBO R82913	N	Y	20	R82150 (20)	Untreated patient	(58)
V 179 D	GTT to GAT	HIV-1 Specific RTI	Trovirdine	Y	?			Found in combination with K103R or Y181C; V179D/Y181C: > 1,000-fold	(43,44)
V 179 D	GTT to GAT	HIV-1 Specific RTI	UC-10 (645129)	Y	?	16			(34)
V 179 E	GTT to GAG	HIV-1 Specific RTI	L-697,661	N	Y	8			(22)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	α-APA R18893 (loviride analog)	Y	?				(59)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	BHAP U-87201E (atevirdine)	N	Y			K103N and Y181C observed with monotherapy	(40)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	BHAP U-88204E	Y	?				(25)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	BHAP U-90152 (delavirdine)	?	Y			K103N/Y181C seen with monotherapy and combination therapy	(45)

Mutations in HIV-1 RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
Y 181 C	TAT to TGT	HIV-1 Specific RTI	BM+51.0836	Y	?				(60)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	DMP 266 (L-743,726)	Y	?	4		L100I/V179D/Y181C: 1,000-fold	(26,27,28)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	E-EBU	Y	?				(50)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	E-EPSeU	Y	?	> 50		Y188C confers greater resistance than Y181C	(61)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	E-EPU	Y	?	> 95		Y188C confers greater resistance than Y181C	(61)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	L-697,593	Y	?	> 100		K103N/Y181C: > 1,000-fold	(46)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	L-697,661	Y	Y	> 30		K103N and Y181C most common with monotherapy	(22,47)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	Loviride (R89439, α-APA)	?	Y				(62)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	Nevirapine	Y	Y	> 100	Other NNRTIs	Can suppress effects of AZT mutations	(23,63,64)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	NSC 648400 (E-BPTU)	Y	?	160	Other NNRTIs		(65)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	TIBO R82913	Y	?	> 100		K103N/Y181C: > 1,000-fold	(33)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	Trovirdine	Y	?		Nevirapine; 9-chloro-TIBO	V179D/Y181C: > 1,000-fold; Found in combination with K103R or V179D	(43,44)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-10 (645129)	Y	?			K101E/Y181C: 200-fold	(35)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-32 (645542)	Y	?	38			(35)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-38 (629243)	Y	?	8-149	Other NNRTIs		(35,66)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-57 (647014)	Y	?			K101E/Y181C: 58-fold	(35)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-68 (638532)	Y	?	5			(35)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-69 (646989)	Y	?			V106A/V181C: 166-fold	(35)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-80 (639475)	Y	?	18			(35)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-81 (615727)	Y	?	53			(34,67)

Mutations in HIV-1 RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-84 (615985)	Y	?	> 118			(35)
Y 181 I	TGT to ATT	HIV-1 Specific RTI	BHAP U-88204E	Y	Y			Appeared after treatment of Y181C-mutated virus with BHAP; high-level resistance to BHAP, nevirapine and TIBO; observed in one nevirapine-treated patient	(68)
Y 181 I	TAT to ATT	HIV-1 Specific RTI	MKC-442 (I-EBU)	Y	N	1,000	All NNRTIs		(56)
Y 181 I	TGT to ATT	HIV-1 Specific RTI	Nevirapine	N	Y	High-level		Observed in one patient	(69)
M 184 I	ATG to ATA	Nucleoside RTI	3TC (lamivudine)	Y	Y			M184V and M184I can suppress effects of AZT resistance mutations	(70,71,72,73)
M 184 T		Nucleoside RTI	3TC (lamivudine)	Y	?			Reduced replication capacity and RT activity	(74)
M 184 V	ATG to GTG	Nucleoside RTI	3TC (lamivudine)	Y	Y	>100	ddI; ddC; (-)-FTC	M184V and M184I can suppress effects of AZT resistance mutations; GTA seen in MT-2 cells in culture	(70,71,72)
M 184 V	ATG to GTG	Nucleoside RTI	ddC	Y	Y	2-5	ddI; 3TC; (-)-FTC		(75)
M 184 V	ATG to GTG	Nucleoside RTI	(-)-FTC	Y	?	> 100	3TC	M184V can suppress effects of AZT mutations	(70,71)
M 184 V	ATG to GTG	Nucleoside RTI	1592U89	Y	N	3	3TC; ddI; ddC	K65R/L74V and/or Y115F with M184V: 10 fold; K65R/M184V: 8-fold; L74V/M184V: 9-fold resistance; L74V/Y115F/M184V: 11-fold	(4,5)
M 184 V	ATG to GTG	Nucleoside RTI	ddI	Y	Y	2-5	ddC; 3TC;(-)-FTC	Rarely observed in patients receiving ddI	(75)
M 184 V	ATG to GTG	Nucleoside RTI	L-FddC	Y	?	> 100	3TC; (-)-FTC		(76)
Y 188 C	TAT to TGT	HIV-1 Specific RTI	E-EPSeU	Y	?	> 250		Y188C is the predominant mutation for E-EPSeU; Y188C confers greater resistance than Y181C	

Mutations in HIV-1 RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
Y 188 C	TAT to TGT	HIV-1 Specific RTI	E-EPU	Y	?	>250		Y188C confers greater resistance than Y181C	(61)
Y 188 C	TAT to TGT	HIV-1 Specific RTI	HEPT	Y	?				(50)
Y 188 C	TAT to TGT	HIV-1 Specific RTI	Nevirapine	N	Y				(29)
Y 188 H	TAT to CAT	HIV-1 Specific RTI	BHAP U-87201E (atevirdine)	N	Y			K101E, Y188H, E233Y and K238T observed with U-87201E/AZT combination therapy	(40)
Y 188 H	TAT to CAT	HIV-1 Specific RTI	TIBO R82913	Y	?				(31)
Y 188 H/L	TAT to CAT/ CTT	HIV-1 Specific RTI	Loviride (R89439, α-APA)	?	Y				(62)
Y 188 L	TAT to TTA	HIV-1 Specific RTI	DMP 266 (L-743,726)	Y	?	1,000			(28)
Y 188 L	TAT to TTA	HIV-1 Specific RTI	TIBO R82913	N	Y				(58)
V 189 I	GTA to ATA	HIV-1 Specific RTI	HBY 097	Y	?	2	Other NNRTIs (2-6)		(15)
G 190 A	GGA to GCA	HIV-1 Specific RTI	Loviride (R89439, α)	?	Y				(77)
G 190 A	GGA to GCA	HIV-1 Specific RTI	Nevirapine	N	Y				(23)
G 190 E	GGA to GAA	HIV-1 Specific RTI	AAP-BHAP (U-104489)	Y	?				(78)
G 190 E	GGA to GAA	HIV-1 Specific RTI	HBY 097	Y	?		Other NNRTIs	Reduces enzymatic activity of RT and viral replication competency	(79)
G 190 E	GGA to GAA	HIV-1 Specific RTI	S-2720	Y	?			Mutation decreases RT activity and viral replication competency	(80)
G 190 E	GGA to GAA	HIV-1 Specific RTI	UC-38 (629243)	Y	N			K101E/G190E: > 100-fold; cross resistance to: TSAO-m^3 T, Nevirapine, TIBO R82913, BHAP U88204; susceptible to L697,661	(34,42)
G 190 Q	GGA to CAA	HIV-1 Specific RTI	HBY 097	Y	?		Other NNRTIs	Appears exclusively in connection with V179D change	(15)

Mutations in HIV-1 RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
G 190 T	GGA to ?	HIV-1 Specific RTI	HBY 097					Appears under lowered drug concentration selection	(51)
H 208 Y	CAT to TAT	Pyrophosphate Analogue RTI	Foscarnet (PFA)	Y	Y	2		Q161L/H208Y: 9-fold PFA resistance; increased susceptibility to AZT (100-fold), nevirapine (20-fold) and TIBO R82150 (30-fold); Q161L/H208Y reverses effects of AZT mutations D67N, K70R, T215Y and K219Q	(18)
L 210 W	TTG to TGG	Nucleoside RTI	AZT	Y	Y			Mutation arises after prolonged AZT therapy in the context of mutations M41L and T215Y	(81,82)
T 215 F	ACC to TTC	Nucleoside RTI	AZT	?	Y			K67N/K70R/T215Y/K219Q: 120-fold	(1,2,3)
T 215 Y	ACC to TAC	Nucleoside RTI	AZT	Y	Y			K67N/K70R/T215Y/K219Q: 120-fold Effect of T215Y is reversed by a ddI mutation (L74V), NNRTI mutations (L100I;Y181C) or (-)-FTC/3TC mutations (M184I/V); M41L/T215Y: 60-70-fold	(1,2,3)
Y 215 C	TTC to TGC	Nucleoside RTI	ddC	N	Y	4		Arises on background of T215Y AZT resistance	(83)
K 219 E	AAA to GAA	Nucleoside RTI	AZT	Y	N				(1,2,3)
K 219 Q	AAA to CAA	Nucleoside RTI	AZT	?	Y			K67N/K70R/T215Y/K219Q: 120-fold	(1,2,3)
E 233 V	GAA to GTA	HIV-1 Specific RTI	BHAP U-87201E (atevirdine)	N	Y			K101E, Y188H, E233Y and K238T observed with U-87201E/AZT combination therapy	(40)
P 236 L	CCT to CTT	HIV-1 Specific RTI	BHAP U-87201E (atevirdine)	Y	N				(84)
P 236 L	CCT to CTT	HIV-1 Specific RTI	BHAP U-90152 (delavirdine)	Y	Y			Sensitizes RT 10-fold to nevirapine, TIBO R82913 and L-697,661	(84)
P 236 L	CCT to CTT	HIV-1 Specific RTI	HEPT	Y	?				(65)
K 238 T	AAA to ACA	HIV-1 Specific RTI	BHAP U-87201E (atevirdine)	N	Y			K101E, Y188H, E233Y and K238T observed with U-87201E/AZT combination therapy	(40)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
R 8 K	CGA to AAA	Protease Inhibitor	A-77003	Y	?	10		R8K/M46I/G48V: 20-fold	(85, 86)
R 8 Q	CGA to CAA	Protease Inhibitor	A-77003	Y	?	10		M46I improves replication competency of R8Q mutant	(85,87)
L 10 F	CTC to TTC	Protease Inhibitor	DMP 450	Y	?			Probably compensatory	(88,89)
L 10 F	CTC to GGC	Protease Inhibitor	VB 11,328	Y	?			L10F/I84V: 8-fold	(90)
L 10 F	CTC to CGC	Protease Inhibitor	VX-478 (141W94)	Y	?				(91)
L 10 F	CTC to CGC	Protease Inhibitor	XM323					L10F/V82A: 2-fold; L10F/K45I/I84V: 50-fold	(92)
L 10 F	CTC to CGC	Protease Inhibitor	SC-55389A	Y	?			L10F/N88S: 10-fold	(93,94)
L 10 I	CTC to ATC	Protease Inhibitor	MK-639 (L-735,524, indinavir)	?	Y				(95)
L 10 I		Protease Inhibitor	Ro 31-8959 (saquinavir)		Y			Found in combination with G48V in vivo	(96)
L 10 R	CTC to CGC	Protease Inhibitor	MK-639 (L-735,524, indinavir)	N	Y			L10R/M46I/L63P/V82T: 4-fold; L10R/M46I/L63P/V82T/I84V: 8-fold	(95)
L 10 V	CTC to GTC	Protease Inhibitor	MK-639 (L-735,524, indinavir)	?	Y				(95)
K 20 M	AAG to ATG	Protease Inhibitor	MK-639 (L-735,524, indinavir)	?	Y				(95)
K 20 R	AAG to AAA	Protease Inhibitor	ABT-538 (ritonavir)	N	Y			K20R/M36I/I54V/V82A: 41-fold	(97,98)
K 20 R	AAG to AAA	Protease Inhibitor	MK-639 (L-735,524, indinavir)	?	Y				(95)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
L 23 I	CTA to ATA	Protease Inhibitor	BILA 2185 BS	Y	?			p1/p6 cleavage site mutation (L to F (CTT to TTT) at P1'); p7(NC)/p1 cleavage site mutation (Q to R (CAG to CGG) at P3, A to V (GCT to GTT) at P2'); L23I/V32I/M46I/I47V/I54M/A71V/I84V:1300-fold	(99)
L 24 I	TTA to ATA	Protease Inhibitor	MK-639 (L-735,524, indinavir)	?	Y				(95)
L 24 V	TTA to GTA	Protease Inhibitor	SC-52151	Y	?	10-20		L24V/G48V/A71V/V75I/P81T: 1000-fold	(94)
D 30 N	GAT to AAT	Protease Inhibitor	AG1343 (nelfinavir)	Y	Y			D30N/A71V: 7-fold; D30N and N88D are most common in vivo after 24 weeks of therapy; they do not cause cross-resistance to other protease inhibitors	(100,101)
V 32 I	GTA to ATA	Protease Inhibitor	A-75925	Y	?	40		V32I and V82I are synergistic mutations yielding 20-fold enzyme resistance	(87,102)
V 32 I	GTA to ATA	Protease Inhibitor	A-77003	Y	?	7 (enzyme resist.)		V32I appears first; progression to V32I/M46V and V32I//M46V/A71V/V82A occurs even in the absence of drug	(87,103)
V 32 I	GTA to ATA	Protease Inhibitor	BILA 1906 BS	Y	?			V32I/A71V: 3-fold; V32I/A71V/M46I/I84V: 500-1,000-fold	(99,104,105)
V 32 I	GTA to ATA	Protease Inhibitor	BILA 2011 (palinavir)	Y	?	1200		Other mutations found in p1/p6 cleavage site	(105)
V 32 I	GTA to ATA	Protease Inhibitor	KNI-272	Y	?	2		V32I/M46I/I84V: 37-fold; V32I/L33F/K45I/F53L/A71V/I84V/L89M: 130-fold	(106)
V 32 I	GTA to ATA	Protease Inhibitor	MK-639 (L-735, 524, indinavir)	Y	Y			V32I/M46L/V82A: 3-fold; V32I/M46L/A71V/V82A: 14-fold	(86)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
L 33 F	TTA to TTC	Protease Inhibitor	ABT-538 (ritonavir)	N	Y				(97)
M 36 I	ATG to ATA	Protease Inhibitor	ABT-538 (ritonavir)	N	Y			M36I/I54V/A71T/V82T: 8-fold; K20R/M36I/I59V/V82A: 41-fold; In vivo, V82 occurs first, often followed by changes at I54, A71 and M36	(97,98)
M 36 I		Protease Inhibitor	AG1343 (nelfinavir)		Y			M46I/L63P/A71V/I84V: 30-fold	(100,101)
K 45 I	AAA to ATA	Protease Inhibitor	XM323					L10F/K45I/I84V: 50-fold	(86)
M 46 F	ATG to TTC	Protease Inhibitor	A-77003	Y	?	4 (enzyme resist.)		Seen with V82A	(87)
M 46 I	ATG to ATA	Protease Inhibitor	A-77003	Y	?			No effect on susceptibility but improves replication competency of R8Q mutant; R8K/M46I/G48V: 20-fold	(85,86)
M 46 I	ATG to ATA	Protease Inhibitor	ABT-538 (ritonavir)	Y	Y			M46I/L63P/A71V/V82F/I84V: 27-fold	(97,98)
M 46 I	ATG to ATA	Protease Inhibitor	AG1343 (nelfinavir)	Y	Y				
M 46 I	ATG to ATA	Protease Inhibitor	BILA 1906 BS	Y	?			V32I/A71V/M46I/I84V: 500-1,000-fold	(96,104,105)
M 46 I	ATG to ATA	Protease Inhibitor	DMP 450	Y	?			Probably compensatory	(88,89)
M 46 I	ATG to ATA	Protease Inhibitor	MK-639 (L-735,524, indinavir)	N	Y			M46I/L63P/V82T: 4-fold; L10R/M46I/L63P/V82T: 4-fold; L10R/M46I/L63P/V82T/I84V: 8-fold	(95,107)
M 46 I	ATG to ATA	Protease Inhibitor	VB 11,328	Y	?			I50V/M46I/I47V: 20-fold	(86,90)
M 46 I	ATG to ATA	Protease Inhibitor	VX-478 (141W94)	Y	?	Nil			
M 46 L	ATG to TTC	Protease Inhibitor	A-77003	Y	?	2-3 (enzyme resist.)			(87)
M 46 L	ATG to TTG	Protease Inhibitor	BILA 1906 BS	Y	?			p1/p6 cleavage site mutation (L to F (CTT to TTT) at P1'); p7(NC)/p1 cleavage site mutation (Q to R (CAG to CGG) at P3, A to V (GCT to GTT) at P2')	(99,104,105)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
M 46 L	ATG to TTG	Protease Inhibitor	MK-639 (L-735,524, indinavir)	Y	Y			V32I/M46L/A71V/V82A: 14-fold; V32I/M46L/V82A: 3-fold	(86)
M 46 L	ATG to CTG	Protease Inhibitor	XM323	Y	?			M46L/V82A: 7-fold; M46L/V82A/L97V: 11-fold	(92)
M 46 V		Protease Inhibitor	A-77003	Y	?			V32I appears first; progression to V32I/M46V and V32I//M46V/A71V/V82A occurs even in the absence of drug	(87,103)
I 47 V	ATA to CTA	Protease Inhibitor	VB 11,328	Y	?			I50V/M46I/I47V: 20-fold	(86,90)
I 47 V	ATA to CTA	Protease Inhibitor	VX-478 (141W94)	Y	?	Nil			(108)
G 48 V	GGG to GTG	Protease Inhibitor	A-77003	Y	?			R8K/M46I/G48V: 20-fold; G48V/I82T: 100-fold	(86)
G 48 V	GGG to GTG	Protease Inhibitor	Ro 31-8959 (saquinavir)	Y	Y			Found in comb. with L10I in vivo; G48V/I84V/L90M: 30-fold; G48V/L90M: >100-fold enzyme resistance; G48V/L90M/I54V: > 50-fold (subtype B or O); No back mutation seen in absence of drug at passage 26	(109,110)
G 48 V	GGG to GTG	Protease Inhibitor	SC-52151	Y	?			G48V/V82A, G48V/L63P/V82A or I54T: 10- to 20-fold; L24V/G48V/A71V/V75I/P81T: 1000-fold	(93,94)
I 50 V	ATT to GTT	Protease Inhibitor	VB 11,328	Y	?	3		I50V/M46I/I47V: 20-fold	(86)
I 50 V	ATT to GTT	Protease Inhibitor	VX-478 (141W94)	Y	?	3			(108)
I 54 L	ATC to ?	Protease Inhibitor	ABT-538 (ritonavir)	N	Y			In vivo, V82 occurs first, often followed by changes at I54, A71 and M36	(97,98)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
I 54 V	ATC to GTC	Protease Inhibitor	ABT-538 (ritonavir)	N	Y			I54V/V82T: 9-fold; K20R/M36I/I54V/V82A: 41-fold; M36I/I54V/A71V/V82T: 8-fold; I54V/A71V/L90M: 7-fold; In vivo, V82 occurs first, often followed by changes at I54, A71 and M36	
I 54 V	ATC to GTC	Protease Inhibitor	MK-639 (L-735,524, indinavir)	?	Y				(95)
I 54 V	ATA to GTA	Protease Inhibitor	Ro 31-8959 (saquinavir)	Y				In subtype O	(109,110)
I 54 V	ATC to GTC	Protease Inhibitor	Ro 31-8959 (saquinavir)	Y				In subtype B	(109,110)
D 60 E	GAT to GAA	Protease Inhibitor	DMP 450	Y	?			Probably compensatory	(81,88)
L 63 P	CTC to CCC	Protease Inhibitor	AG1343 (nelfinavir)		Y			M46I/L63P/A71V/I84V: 30-fold	(100,101)
L 63 P	CTC to CCC	Protease Inhibitor	BILA 2011 (palinavir)	Y	?				
L 63 P	CTC to CCC	Protease Inhibitor	MK-639 (L-735,524, indinavir)	N	Y			M46I/L63P/V82T: 4-fold; L10R/M46I/L63P/V82T/I84V: 8-fold; L10R/M46I/L63P/V82T: 4-fold	(95,107)
A 71 T	GCT to ACT	Protease Inhibitor	BMS 186,318	Y	?			A71T/V82A: 15-fold	(111,112)
A 71 T	GCT to ACT	Protease Inhibitor	MK-639 (L-735,524, indinavir)	?	Y				(95)
A 71 V		Protease Inhibitor	A-77003	Y	?			V32I appears first; progression to V32I/M46V and V32I//M46V/A71V/V82A occurs even in the absence of drug; M46I/L63P/A71V/V82F/I84V: 27-fold	(87,103)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
A 71 V	GCT to GTT	Protease Inhibitor	ABT-538 (ritonavir)	Y	Y			K20R/M36I/I54V/A71V/V82T: 28-fold; M36I/I54V/A71V/V82T: 8-fold; I54V/A71V/I/V82A/L90M: 7-fold; In vivo, V82 occurs first, often followed by changes at I54, A71 and M36	
A 71 V	GCT to GTT	Protease Inhibitor	AG1343 (nelfinavir)	Y	?	5		D30N/A71V: 7-fold; M46I/L63P/A71V/I84V: 30-fold	(100,101)
A 71 V	GCT to GTT	Protease Inhibitor	BILA 1906 BS	Y	?			V32I/A71V: 3-fold; V32I/A71V/M46I/I84V: 500-1,000-fold	(99,104,105)
A 71 V	GCT to GTT	Protease Inhibitor	BILA 2011 (palinavir)	Y	?				
A 71 V	GCT to GTT	Protease Inhibitor	MK-639 (L-735,524, indinavir)	Y	Y			V32I/M46L/A71V/V82A: 14-fold	(86)
A 71 V	GCT to GTT	Protease Inhibitor	SC-52151	Y	?			A71V/V75I/P81T: 20- to 30-fold; L24V/G48V/A71V/V75I/P81T: 1000-fold; N88D or I11V/M46I/F53L/A71V/N88D:10- to 20-fold	(93,94)
V 75 I	GTA to ATA	Protease Inhibitor	SC-52151	Y	?			L24V/G48V/A71V/V75I/P81T: 1000-fold; A71V/V75I/P81T: 20- to 30-fold; L24V/G48V/A71V/V75I/P81T: 1000-fold	(93,94)
V 77 I		Protease Inhibitor	AG1343 (nelfinavir)	Y	Y				
P 81 T	CCT to ACT	Protease Inhibitor	SC-52151	Y	?			A71V/V75I/P81T: 20- to 30-fold; L24V/G48V/A71V/V75I/P81T: 1000-fold	(93,94)
I 82 T	ATC to ACC	Protease Inhibitor	A-77003	Y	?			G48V/I82T: 100-fold (82T was derived from in vitro passage of 82I)	(111)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
V 82 A	GTC to GCC	Protease Inhibitor	A-77003	Y	?			Rare; seen with M46F; V32I appears first; progression to V32I/M46V and V32I//M46V/A71V/V82A occurs even in the absence of drug	(87,103,111)
V 82 A	GTC to GCC	Protease Inhibitor	ABT-538 (ritonavir)	N	Y	2		In vivo, V82 occurs first, often followed by changes at I54, A71 and M36	(97,98)
V 82 A	GTC to GCC	Protease Inhibitor	BMS 186,318	Y	?			A71T/V82A: 15-fold	(112,113)
V 82 A	GTC to GCC	Protease Inhibitor	MK-639 (L-735,524, indinavir)	Y	Y			V32I/M46L/V82A: 3-fold; V32I/M46L/A71V/V82A: 14-fold	(86,95,107,114)
V 82 A	GTC to GCC	Protease Inhibitor	P9941	Y	?	6-8			(115)
V 82 A	GTC to GCC	Protease Inhibitor	SC-52151	Y	?			G48V/V82A, G48V/L63P/V82A or I54T: 10- to 20-fold	(93,94)
V 82 A		Protease Inhibitor	SKF108922	Y	?				
V 82 A	GTC to GCC	Protease Inhibitor	XM323	Y	?			V82A/M46L: 7-fold; V82A/M46L/L97V: 11-fold; L10F/V82A: 2-fold; V82A/L97V: 3-fold	(92)
V 82 F	GTC to TTC	Protease Inhibitor	ABT-538 (ritonavir)	Y	Y			V82F/I84V: 8- to 10-fold; M46I/L63P/A71V/V82F/I84V: 27-fold	(97,98)
V 82 F	GTC to TTC	Protease Inhibitor	MK-639 (L-735,524, indinavir)	?	Y				(95,107)
V 82 F	GTC to TTC	Protease Inhibitor	XM323	Y	?			V82F/I84V: 92-fold	(92)
V 82 I	GTC to ATC	Protease Inhibitor	A-77003	Y	?			No resistance alone but V32I and V82I are synergistic mutations yielding 20-fold enzyme resistance (82T was derived from in vitro passage of 82I)	(87,111)
V 82 I	GTC to ATC	Protease Inhibitor	XM323	Y	?	< 2			(92)
V 82 S	GTC to TCC	Protease Inhibitor	ABT-538 (ritonavir)	N	Y	6		In vivo, V82 occurs first, often followed by changes at I54, A71 and M36	(97,98)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
V 82 T	GTC to ACC	Protease Inhibitor	ABT-538 (ritonavir)	N	Y	3		In vivo, V82 occurs first, often followed by changes at I54, A71 and M36; V82T has reduced replication efficacy in natural background; M36I/I54V/A71V/V82T: 8-fold; I54V/V82T: 9-fold	
V 82 T	GTC to ACC	Protease Inhibitor	MK-639 (L-735,524, indinavir)	N	Y			M46I/L63P/V82T: 4-fold; L10R/M46I/L63P/V82T: 4-fold; L10R/M46I/L63P/V82T/I84V: 8-fold	(95,107)
V 82 T		Protease Inhibitor	SKF108842	Y	?				(116)
I 84 A	ATA to GCA	Protease Inhibitor	BILA 1906 BS	Y	?				(99,104,105)
I 84 A	ATG to ATA	Protease Inhibitor	BILA 2011 (palinavir)	Y	?			I84A is the most common mutation	(99,104,105)
I 84 V	ATA to GTA	Protease Inhibitor	ABT-538 (ritonavir)	Y	Y			M46I/L63P/A71V/V82F/I84V: 27-fold; V82F/I84V: 8- to 10-fold; M46I/L63P/A71V/V82F/I84V: 27-fold	(97,98)
I 84 V	ATA to GTA	Protease Inhibitor	AG1343 (nelfinavir)		?			M46I/L63P/A71V/I84V: 30-fold	(100,101)
I 84 V	ATA to GTA	Protease Inhibitor	BILA 1906 BS	Y	?			V32I/A71V/M46I/I84V: 500-1,000-fold	(99,104,105)
I 84 V	ATA to GTA	Protease Inhibitor	DMP 450	Y	?			S1 subsite	(88,89)
I 84 V	ATA to CTA	Protease Inhibitor	MK-639 (L-735,524, indinavir)	N	Y			G48V/I84V/L90M: 30-fold; L10R/M46I/L63P/V82T/I84V: 8-fold	(95,107)
I 84 V	ATA to GTA	Protease Inhibitor	Ro 31-8959 (saquinavir)	Y	?				(86)
I 84 V	ATA to GTA	Protease Inhibitor	RPI-312	Y	?	5			(117)
I 84 V		Protease Inhibitor	SKF108842	Y	?				(116)
I 84 V	ATA to GTA	Protease Inhibitor	VB 11,328	Y	?			L10F/I84V: 8-fold	(86,90)
I 84 V	ATA to GTA	Protease Inhibitor	VX-478 (141W94)	Y	?				(108)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
I 84 V	ATA to GTA	Protease Inhibitor	XM323	Y	?	12		V82F/I84V: 92-fold; L10F/K45I/I84V: 50-fold	(86,92)
N 88 D		Protease Inhibitor	AG1343 (nelfinavir)	Y	Y			D30N and N88D are most common in vivo after 24 weeks of therapy; they do not cause cross-resistance to other protease inhibitors	(100,101)
N 88 D	AAT to GAT	Protease Inhibitor	SC-52151	Y	?			N88D or I11V/M46I/F53L/A71V/N88D:10- to 20-fold	(93,94)
N 88 S	AAT to AGT	Protease Inhibitor	SC-55389A	Y	?	20		N88S/L10F: 10-fold	(93,94)
L 90 M	TTG to ATG	Protease Inhibitor	ABT-538 (ritonavir)	N	Y			82A/54V/I/71V/90L/M: 7-fold	(97,98)
L 90 M	TTG to ATG	Protease Inhibitor	AG1343 (nelfinavir)	N	Y			Rare in patients	(100,101)
L 90 M	TTG to ATG	Protease Inhibitor	MK-639 (L-735,524, indinavir)	?	Y				(95,107)
L 90 M	TTG to ATG	Protease Inhibitor	Ro 31-8959 (saquinavir)	Y	Y			G48V/L90M: >100-fold enzyme resistance; double mutant rare in vivo; L90M most common in vivo; G48V/I84V/L90M: 30-fold	(109)
L 97 V	TTA to GTA	Protease Inhibitor	XM323	Y	?			No resistance alone; V82A/L97V: 3-fold; V82A/M46L/L97V: 11-fold	(92)

Mutations in Envelope that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
S 113 N	AGT to AAT	Fusion/Binding Inhibitor	Dextran sulphate (DS)	Y	?			S113N/S134N/K269E/Q278E/N293D/N323S/R387I: 250-fold; 113 is in the V1 loop region	(118,119)
S 134 N	AGC to AAC	Fusion/Binding Inhibitor	Dextran sulphate (DS)	Y	?			V2 loop region; S113N/S134N/K269E/Q278E/N293D/N323S/R387I: 250-fold	(118,119)
F 145 L	TTC to TTA	Fusion/Binding Inhibitor	JM-3100	Y	?			Combination of mutations: 2- to 100-fold	(120,121)
N 188 K	AAT to AAA	Fusion/Binding Inhibitor	Siamycin I	Y	?			N188K/G332E/N351D/A550T/N633D/L762S: 9-fold	(122)
I 228 V	ATA to GTA	Fusion/Binding Inhibitor	JM-2763	Y	?			Combination of mutations	
K 269 E	AAA to GAA	Fusion/Binding Inhibitor	Dextran sulphate (DS)	Y	?			V3 loop region; S113N/S134N/K269E/Q278E/N293D/N323S/R387I: 250-fold	(118,119)
N 270 S	AAT to AGT	Fusion/Binding Inhibitor	JM-3100	Y	?				
R 272 T	AGA to ACA	Fusion/Binding Inhibitor	JM-3100	Y	?				
S 274 R	AGT to AGA	Fusion/Binding Inhibitor	JM-2763	Y	?			Combination of mutations: 95- to 792-fold	(120,121)
S 274 R	AGT to AGA	Fusion/Binding Inhibitor	JM-3100	Y	?	DS (> 7 to 6,667)			
Q 278 H	CAG to CAT	Fusion/Binding Inhibitor	Dextran sulphate (DS)	Y	?			V3 loop region; S113N/S134N/K269E/Q278E/N293D/N323S/R387I: 250-fold	(118,119)
Q 278 H	CAG to CAT	Fusion/Binding Inhibitor	JM-2763	Y	?				
Q 278 H	CAG to CAC	Fusion/Binding Inhibitor	JM-3100	Y	?				
I 288 V	ATA to GTA	Fusion/Binding Inhibitor	JM-3100	Y	?				
N 293 D	AAT to GAT	Fusion/Binding Inhibitor	Dextran sulphate (DS)	Y	?			V3 loop region; S113N/S134N/K269E/Q278E/N293D/N323S/R387I: 250-fold	(118,119)
N 293 H	AAT to CAT	Fusion/Binding Inhibitor	JM-3100	Y	?				
A 297 T	GCA to ACA	Fusion/Binding Inhibitor	JM-2763	Y	?				
A 297 T	GCA to ACA	Fusion/Binding Inhibitor	JM-3100	Y	?				
N 323 S	AAT to AGT	Fusion/Binding Inhibitor	Dextran sulphate (DS)	Y	?			C3 region; S113N/S134N/K269E/Q278E/N293D/N323S/R387I: 250-fold	(118,119)

Mutations in Envelope that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	Fold-resistance	Cross-resistance (-fold)	Comments	Refs
G 332 E	GGA to GAA	Fusion/Binding Inhibitor	Siamycin I	Y	?			N188K/G332E/N351D/A550T/N633D/ L762S: 9-fold	(122)
N 351 D	AAT to GAT	Fusion/Binding Inhibitor	Siamycin I	Y	?			N188K/G332E/N351D/A550T/N633D/ L762S: 9-fold	(122)
P 385 L	CCA to CTA	Fusion/Binding Inhibitor	JM-2763	Y	?				
P 385 L	CCA to CTA	Fusion/Binding Inhibitor	JM-3100	Y	?				
R 387 I	AGA to ACA	Fusion/Binding Inhibitor	Dextran sulphate (DS)	Y	?			CD4 binding region; S113N/S134N/ K269E/Q278E/N293D/N323S/R387I: 250-fold	(118,119)
Q 410 E	CAA to GAA	Fusion/Binding Inhibitor	JM-3100	Y	?				
S 433 P	TCC to CCC	Fusion/Binding Inhibitor	JM-3100	Y	?				
V 457 I	GTA to ATA	Fusion/Binding Inhibitor	JM-3100	Y	?				
A 550 T	GCC to ACC	Fusion/Binding Inhibitor	Siamycin I	Y	?			N188K/G332E/N351D/A550T/N633D/ L762S: 9-fold	(122)
N 633 D	AAT to GAT	Fusion/Binding Inhibitor	Siamycin I	Y	?			N188K/G332E/N351D/A550T/N633D/ L762S: 9-fold	(122)
L 762 S	TTG to TCG	Fusion/Binding Inhibitor	Siamycin I	Y	?			N188K/G332E/N351D/A550T/N633D/ L762S: 9-fold	(122)

Mutations that confer multiple drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
A 62 V	GCC to GTC	Multiple Drug Resistance	AZT+ ddI/ddC	N	Y	Nil		Associated with 75, 77, 116 & 151; A62V/V75I/F77L/F116Y/Q151M: AZT 190-fold	(123,124)
V 75 I	GTA to ATA	Multiple Drug Resistance	AZT+ ddI/ddC	N	Y	Nil		Associated with 77, 116 & 151; A62V/V75I/F77L/F116Y/Q151M: AZT 190-fold	(123,124)
F 77 L	TTC to CTC	Multiple Drug Resistance	AZT+ ddI/ddC	N	Y	Nil		Associated with 75, 116 & 151; A62V/V75I/F77L/F116Y/Q151M: AZT 190-fold	(123,124)
F 116 Y	TTT to TAT	Multiple Drug Resistance	AZT+ ddI/ddC	N	Y	Nil		Associated with 75, 77 & 151; A62V/V75I/F77L/F116Y/Q151M: AZT 190-fold	(123,124)
Q 151 M	CAG to ATG	Multiple Drug Resistance	AZT+ ddI/ddC	N	Y	AZT: 10; ddI/ddC: 5		Pivotal MDR mutation (first to occur and is then found in association with various of the other four mutations); A62V/V75I/F77L/F116Y/Q151M: AZT 190-fold; ddI 50-fold; ddC 20-fold; d4T > 10-fold	(123,124,125)

Abbreviations

Amino acids

A	alanine
C	cysteine
D	aspartate
E	glutamate
F	phenyalanine
G	glycine
H	histidine
I	isoleucine
K	lysine
L	leucine
M	methionine
N	asparagine
P	proline
Q	glutamine
R	arginine
S	serine
T	threonine
V	valine
W	tryptophan
Y	tyrosine

Compounds

1592U89	(1*S*,4*R*)-4-[2-amino-6-cyclopropyl-amino)-9*H*-purin-9-yl]-2-cyclopentene-1-methanol succinate (a carbovir analogue, Glaxo Wellcome)
3TC	(-)-β-L-2′,3′-dideoxy-3′-thiacytidine (Glaxo Wellcome)
α-APA R18893	α-nitro-anilino-phenylacetamide
A-77003, A-75925 and A-80987	C2 symmetry-based protease inhibitors (Abbott Laboratories)
AAP-BHAP	bisheteroarylpiperazine analogue (Pharmacia & Upjohn)
ABT-538	C2 symmetry-based protease inhibitor (Abbott Laboratories)
AZdU	3′-azido-2′,3′-dideoxyuridine
AZT	3′-azido-3′-deoxythymidine (Glaxo Wellcome)
AZT-p-ddI	3′-azido-3′-deoxythymidilyl-(5′,5′)-2′,3′-dideoxyinosinic acid (Ivax)
BHAP	bisheteroarylpiperazine
BILA 1906	*N*-{1*S*-[[[3-[2*S*-(1,1-dimethylethyl)amino]carbonyl-4*R*-]3-pyridinylmethyl)thio]-1-piperidinyl]-2*R*-hydroxy-1*S*-(phenylmethyl)propyl]amino]carbonyl]-2-methylpropyl}-2-quinolinecarboxamide (Bio-Mega/Boehringer Ingelheim)
BILA 2185	*N*-(1,1-dimethylethyl)-1-[2*S*-[[2-2,6-dimethyphenoxy)-1-oxoethyl]amino]-2*R*-hydroxy-4-phenylbutyl]4*R*-pyridinylthio)-2-piperidine-carboxamide (Bio-Mega/Boehringer Ingelheim)
BM+51.0836	thiazolo-isoindolinone derivative
BMS 186,318	aminodiol derivative HIV-1 protease inhibitor (Bristol-Myers Squibb)

Abbreviations (cont)

Compounds (cont)

d4API	9-[2,5-dihydro-5-(phosphonomethoxy)-2-furanel]adenine (Gilead Sciences)
d4C	2′,3′-didehydro-2′,3′-dideoxycytidine
d4T	2′,3′-didehydro-3′-deoxythymidine (Bristol-Myers Squibb)
ddC	2′,3′-dideoxycytidine (Roche)
ddI	2′,3′-dideoxyinosine (Bristol-Myers Squibb)
DMP 266	a 1,4-dihydro-2*H*-3,1-benzoxazin-2-one
DMP 450	[4*R*-(4-α,5-α,6-β,7-β)]-hexahydro-5,6-bis(hydroxy)-1,3-bis(3-amino)phenyl]methyl)-4,7-bis(phenylmethyl)-2*H*-1,3-diazepin-2-one-bismesylate (Avid Therapeutics)
DXG	(-)-β-D-dioxolane-guanosine
EBU-dM	5-ethyl-1-ethoxymethyl-6-(3,5-dimethylbenzyl)uracil
E-EBU	5-ethyl-1-ethoxymethyl-6-benzyluracil
DS	dextran sulphate
E-EPSeU	1-(ethoxymethyl)-(6-phenylselenyl)-5-ethyluracil
E-EPU	1-(ethoxymethyl)-(6-phenyl-thio)-5-ethyluracil
(-)-FTC	(-)-β-L-2′,3′-dideoxy-5-fluoro-3′-thiacytidine (Triangle Pharmaceuticals)
HBY 097	(*S*)-4-isopropoxycarbonyl-6-methoxy-3-(methylthio-methyl)-3,4-dihydroquinoxalin-2(1*H*)-thione
HEPT	1-[(2-hydroxyethoxy)methyl]6-(phenylthio)thymine
JM2763	1,1′-(1,3-propanediyl)-bis-1,4,8,11-tetraazacyclo-tetradecane (Johnson Matthey)
JM3100	1,1′-[1,4-phenylenebis-(methylene)]bis-(1,4,8,11-tetraazacyclotetradecane) octahydrochloride dihydrate (Johnson Matthey)
KNI-272	(2*S*,3*S*)-3-amino-2-hydroxy-4-phenylbutyric acid-containing tripeptide
L-697,593	5-ethyl-6-methyl-3-(2-phthalimido-ethyl)pyridin-2(1*H*)-one
L-697,661	3-[(-4,7-dichloro-1,3-benzoxazol-2-yl)methyl]amino-5-ethyl-6-methylpyridin-2(1*H*)-one
L-FDDC	(-)-β-L-5-fluoro-2′,3′-dideoxy-cytidine
L-FDOC	(-)-β-L-5-fluoro-dioxolane cytosine
MK-639	hydroxy-aminopentane amide HIV-1 protease inhibitor (Merck & Co)
MKC442	6-benzyl-1-ethoxymethyl-5-isopropyluracil (I-EBU, Triangle Pharmaceuticals/Mitsubishi)
nevirapine	11-cyclopropyl-5,11-dihydro-4-methyl-6*H*-dipyridol[3,2-b:2′,3′-e] diazepin-6-one (Boehringer Ingelheim)
NNRTI	non-nucleoside reverse transcriptase inhibitor
NSC648400	1-benzyloxymethyl-5-ethyl-6-(alpha-pyridylthio)uracil (E-BPTU)
P9941	[2-pyridylacetyl-IIePheAla-y(CHOH)]$_2$ (Dupont Merck)
PFA	phosphonoformate (foscarnet, Astra)
PMEA	9-(2 phosphonylmethoxyethyl)adenine (Gilead Sciences)
PMPA	(*R*)-9-(2-phosphonyl-methoxypropyl)adenine (Gilead Sciences)
Ro 31-8959	hydroxyethylamine derivative HIV-1 protease inhibitor (Roche)

Abbreviations (cont)

Compounds (cont)

RPI-312	1-[(3*S*)-3-(n-alpha-benzyloxycarbonyl)-l-asparginyl)-amino-2-hydroxy-4-phenyl-butyryl]-*n*-tert-butyl-l-proline amide (peptidyl protease inhibitor)
RT	reverse transcriptase
S-2720	6-chloro-3,3-dimethyl-4-(isopropenyl-oxycarbonyl)-3,4-dihydro-quinoxalin-2(1*H*)thione
SC-52151	hydroxyethylurea isostere protease inhibitor (Searle)
SC-55389A	hydroxyethyl-urea isostere protease inhibitor (Searle)
TIBO R82150	(+)-(5*S*)-4,5,6,7-tetrahydro-5-methyl-6-(3-methyl-2-butenyl)-imidazo[4,5,1-*jk*][1,4]-benzodiazepin-2(1*H*)-thione (Janssen)
TIBO 82913	(+)-(5*S*)-4,5,6,7,-tetrahydro-9-chloro-5-methyl-6-(3-methyl-2-butenyl)-imidazo-[4,5,1-*jk*]-[1,4]benzo-diazepin-2(1it H)-thione (Janssen)
TSAO-m^3T	[2′,5′-bis-*O*-(tert-butyl-dimethylsilyl)-3′-spiro-5′-(4′-amino-1′,2′-oxathiole-2′,2′-dioxide)]-β-D-pentofuranosyl-N^3-methylthymine
U-90152	1-[3-[(1-methylethyl)-amino]-2-pyridinyl]-4-[[5-[(methylsulphonyl)-amino]-1*H*-indol-2yl]carbonyl]-piperazine
UC	thiocarboxanilide derivatives (Uniroyal Chemical Co)
UC-781	*N*-[4-chloro-3-(3-methyl-2-butenyloxy)phenyl]-2-methyl-3-furan-carbothioamide
UC-82	*N*-[4-chloro-3-(3-methyl-2-butenyloxy)phenyl]-2-methyl-3-thiophene-carbothioamide
VB 11,328	hydroxyethyl-sulphonamide protease inhibitor (Vertex Pharmaceuticals)
VX-478	hydroxyethylsulphonamide protease inhibitor (Vertex Pharmaceuticals)
XM 323	cyclic urea protease inhibitor (Dupont Merck)

References

[1] Larder BA and Kemp SD. Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science* 1989; **246**:1155–1158.

[2] Larder BA, Coates KE and Kemp SD. Zidovudine-resistant human immunodeficiency virus selected by passage in cell culture. *Journal of Virology* 1991; **65**(10):5232–5236.

[3] Kellam P, Boucher CA and Larder BA. Fifth mutation in human immunodeficiency virus type 1 reverse transcriptase contributes to the development of high-level resistance to zidovudine. *Proceedings of the National Academy of Sciences, USA* 1992; **89**(5):1934–1938.

[4] Tisdale M, Parry NR, Cousens D, St Clair MH and Boone LR. Anti-HIV activity of (lS,4R)-[2-amino-6-cyclopropylamino)-9H-purin-9-yl]-2-cyclopentene-1-methanol (1592U89). 34th Interscience Conference on Antimicrobial Agents and Chemotherapy, Orlando, FL, USA, 2–5 October, 1996. Abstract 92.

[5] Harrigan PR, Tisdale M, Najera I, Cousens D, St Clair M, Stone C, Kohli A, Myers R and Larder BA. Antiretroviral activity and resistance to 1592U89, a novel HIV RT inhibitor. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3–6 July, 1996. Abstract 16.

[6] Zhang D, Caliendo AM, Eron JJ, DeVore KM, Kaplan JC, Hirsch MS and D'Aquila RT. Resistance to 2′,3′-dideoxycytidine conferred by a mutation in codon 65 of the human immunodeficiency virus type 1 reverse transcriptase. *Antimicrobial Agents and Chemotherapy* 1994; **38**(2):282–287.

[7] Gu Z, Gao Q, Fang H, Salomon H, Parniak MA, Goldberg E, Cameron JM and Wainberg MA. Identification of a mutation at codon 65 in the IKKK motif of reverse transcriptase that encodes human immunodeficiency virus resistance to 2′,3′-dideoxycytidine and 2′,3′-dideoxy-3′-thiacytidine. *Antimicrobial Agents and Chemotherapy* 1994; **38**(2):275–281.

[8] Schinazi RF, Boudinot FD, Manouilov KK, Mellors JW, McMillan A, Schlueter-Wirtz S, Lloyd R Jr, Korba BE, Tennant B and Chu CK. Anti-HBV and anti-HIV activities of dioxolane-purine nucleosides. Ninth International Conference on Antiviral Research, Urabandai, Fukushima, Japan, 19–24 May 1996. Abstract 20.

[9] Mellors JW, Bazmi H, Chu CK and Schinazi RF. K65R mutation in HIV-1 reverse transcriptase causes resistance to (-)-β-D-dioxolane-guanine and reverses AZT resistance. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3–6 July, 1996. Abstract 7.

[10] Foli A, Sogocio KM, Anderson B, Kavlick M, Saville MW, Wainberg MA, Gu X, Cherrington J, Mitsuya H and Yarchoan R. In vitro selection and molecular characterization of human immunodeficiency virus type 1 with reduced sensitivity to 9-[2-(phosphonomethoxy-ethyl]adenine (PMEA). *Antiviral Research* 1996; (in press).

[11] Gu Z, Salomon H, Cherrington JM, Mulato AS, Chen MS, Yarchoan R, Foli A, Sogocio KM and Wainberg MA. K65R mutation of human immunodeficiency virus type 1 reverse transcriptase encodes cross-resistance to 9-(2-phosphonylmethoxyethyl)adenine. *Antimicrobial Agents and Chemotherapy* 1995; **39**(8):1888–1891.

[12] Fitzgibbon JE, Howell RM, Haberzettl CA, Sperber SJ, Gockle DJ and Dubin DT. Human immunodeficiency virus type 1 pol gene mutations which cause decreased susceptibility to 2′,3′-dideoxycytidine. *Antimicrobial Agents and Chemotherapy* 1992; **36**:153–157.

[13] Cherrington J, Mulato AS, Fuller MD and Chen MS. A novel mutation (K70E) in HIV-1 reverse transcriptase confers decreased susceptibility to 9-[2-(phosphonomethoxy)ethyl]adenine (PMEA) in vitro. *Antimicrobial Agents and Chemotherapy* 1996; (in press).

[14] St Clair MH, Martin JL, Tudor-Williams G, Bach MC, Vavro CL, King DM, Kellam P, Kemp SD and Larder BA. Resistance to ddI and sensitivity to AZT induced by a mutation in HIV-1 reverse transcriptase. *Science* 1991; **253**:1557–1559.

[15] Kleim JP, Rosner M, Winkler I, Paessens A, Kirsch R, Hsiou Y, Arnold E and Riess G. Selective pressure of a quinoxaline nonnucleoside inhibitor of human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT) on HIV-1 replication results in the emergence of nucleoside RT-inhibitor-specific (RT Leu-74 to Val or Ile and Val-75 to Leu or Ile) HIV-1 mutants. *Proceedings of the National Academy of Sciences, USA* 1996; **93**(1):34–38.

[16] Lacey SF and Larder BA. Novel mutation (V75T) in human immunodeficiency virus type 1 reverse transcriptase confers resistance to 2′,3′-didehydro-2′,3′-dideoxythymidine in cell culture. *Antimicrobial Agents and Chemotherapy* 1994; **38**(6):1428–1432.

[17] Schinazi RF, Stuyver L, Wyseur A, Lloyd RM Jr, Hough L, Rombout A, Rossau R, and Rimland D. Proviral and plasma virus genotyping using a line probe assay in nucleoside treated HIV infected Veterans Affairs patients. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3–6 July, 1996. Abstract 65.

[18] Mellors J, Bazmi H, Schinazi RF, Roy B, Hsiou Y, Arnold E, Weir J and Mayers D. Novel mutations in the reverse transcriptase of human immunodeficiency virus type 1 reduce susceptibility to foscarnet in laboratory and clinical isolates. *Antimicrobial Agents and Chemotherapy* 1995; **39**(5):1087–1092.

[19] Tachedjian G, Hooker DJ, Gurusinghe AD, Bazmi H, Deacon NJ, Mellors J, Birch C and Mills J. Characterisation of foscarnet-resistant strains of human immunodeficiency virus type 1. *Virology* 1995; **212**(1):58–68.

[20] Tachedjian G, Mellors J, Bazmi H, Birch C and Mills J. Zidovudine resistance is suppressed by mutations conferring resistance of human immunodeficiency virus type 1 to foscarnet. *Journal of Virology* 1996; **70**:7171–7181.

[21] Prasad VR, Lowy I, de los Santos T, Chiang L and Goff SP. Isolation and characterization of a dideoxyguanosine triphosphate-resistant mutant of human immunodeficiency virus reverse transcriptase. *Proceedings of the National Academy of Sciences, USA* 1991; **88**(24):11363–11367.

[22] Byrnes VW, Sardana VV, Schleif WA, Condra JH, Waterbury JA, Wolfgang JA, Long WJ, Schneider CL, Schlabach AJ, Wolanski BS, Graham DJ, Gotlib L, Rhodes A, Titus DL, Roth E, Blahy OM, Quintero JC, Staszewski S and Emini EA. Comprehensive mutant enzyme and viral variant assessment of human immunodeficiency virus type 1 reverse transcriptase resistance to nonnucleoside inhibitors. *Antimicrobial Agents and Chemotherapy* 1993; **37**(8):1576–1579.

[23] Richman DD, Havlir D, Corbeil J, Looney D, Ignacio C, Spector SA, Sullivan J, Cheeseman S, Barringer K, Pauletti D, Shih CK, Myers M and Griffin J. Nevirapine resistance mutations of human immunodeficiency virus type 1 selected during therapy. *Journal of Virology* 1994; **68**(3):1660–1666.

[24] Balzarini J, Karlsson A, Perez-Perez MJ, Camarasa MJ, Tarpley WG and De Clercq E. Treatment of human immunodeficiency virus type 1 (HIV-1)-infected cells by combinations of HIV-1-specific inhibitors results in a different resistance pattern than does treatment with single-drug therapy. *Journal of Virology* 1993; **67**(9):5353–5359.

[25] Vasudevachari MB, Battista C, Lane HC, Psallidopoulos MC, Zhao B, Cook J, Palmer JR, Romero DL, Tarpley WG and Salzman NP. Prevention of the spread of HIV-1 infection with nonnucleoside reverse transcriptase inhibitors. *Virology* 1992; **190**(1):269–2977.

[26] Winslow DL, Garber S, Reid C, Scarnati H, Korant B, Emini E and Anton ED. Development of high-level resistance to DMP 266 requires multiple mutations in the reverse transcriptase gene. Fourth International Workshop on HIV Drug Resistance, Sardinia, Italy, 6–9 July, 1995. Abstract 13.

[27] Winslow DL, Reid C, Garber S, Scarnati H, Rayner M and Anton E. Selection conditions affect the evolution of specific mutations in the reverse transcriptase gene associated with resistance to DMP 266. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3–6 July, 1996. Abstract 10.

[28] Young SD, Britcher SF, Tran LO, Payne LS, Lumma WC, Lyle TA, Anderson JR, Huff PS, Olsen DB, Carroll SS, Pettibone DJ, O'Brien JA, Ball RG, Balani SK, Lin JH, Chen IW, Schleif WA, Sardana VV, Long WJ, Byrenes VW and Emini EA. L-743,726 (DMP-266): a novel, highly potent nonnucleoside inhibitor of the human immunodeficiency virus type 1 reverse transcriptase. *Antimicrobial Agents and Chemotherapy* 1995; **39**(12):2602–2609.

[29] Richman DD. Resistance of clinical isolates of human immunodeficiency virus to antiretroviral agents. *Antimicrobial Agents and Chemotherapy* 1993; **37**(6):1207–1213.

[30] Mellors JW, Im GJ, Tramontano E, Winkler SR, Medina DJ, Dutschman GE, Bazmi HZ, Piras G, Gonzalez CJ and Cheng YC. A single conservative amino acid substitution in the reverse transcriptase of human immunodeficiency virus-1 reverse transcriptase confers resistance to TIBO R82150. *Molecular Pharmacology* 1993; **43**(1):11–16.

[31] Balzarini J, Karlsson A, Perez-Perez MJ, Vrang L, Walbers J, Zhang H, Oberg B, Vandamme AM, Camarasa MJ and De Clercq E. HIV-1 specific reverse transcriptase inhibitors show differential activity against HIV-1 mutant strains containing different amino acid substitutions in the reverse transcriptase. *Virology* 1993; **192**:246–253.

[32] Byrnes V, Blahy O, Condra J, Gotlib L, Graham D, Long W, Quintero J, Rhodes A, Roth E, Sardana V, Schlabach A, Schleif W, Schneider C, Titus D, Wolanski B, Wolfgang J and Emini E. Phenotypic susceptibility of human immunodeficiency virus type 1 RT containing substitutions which engender resistance to nucleoside and non-nucleoside inhibitors. Third Workshop on Viral Resistance, 1993, Gaithersburg, MD, USA.

[33] Larder, BA. 3′-Azido-3′-deoxythymidine resistance suppressed by a mutation conferring human immunodeficiency virus type 1 resistance to nonnucleoside reverse transcriptase inhibitors. *Antimicrobial Agents and Chemotherapy* 1992; **36**(12):2664–2669.

[34] Balzarini J, Perez-Perez MJ, Velazquez S, San-Felix A, Camarasa MJ, De Clercq E and Karlsson A. Suppression of the breakthrough of human immunodeficiency virus type 1 (HIV-1) in cell culture by thiocarboxanilide derivatives when used individually or in combination with other HIV-1-specific inhibitors (ie, TSAO derivatives). *Proceedings of the National Academy of Sciences, USA* 1995; **92**:5470–5474.

[35] Buckheit RW, Kinjerski TL, Fliakas-Boltz V, Russell JD, Stup TL, Pallansch LA, Brouwer WG, Dao DC, Harrison WA, Schultz RJ, Bader JP and Yang SS. Structure-activity and cross-resistance evaluations of a series of human immunodeficiency virus type 1-specific compounds related to oxathiin carboxanilide. *Antimicrobial Agents and Chemotherapy* 1995; **39**(12):2718–2727.

[36] Balzarini J, Pelemans H, Aquaro S, Perno CF, Witvrouw M, Schols D, De Clercq E and Karlsson A. Highly favourable antiviral activity and resistance profile of the novel thiocarboxanilide pentenyloxy ether derivatives UC-781 and UC-82 as inhibitors of human immunodeficiency virus type 1 (HIV-1) replication. *Molecular Pharmacology* 1996; (in press).

[37] Balzarini J, Brouwer WG, Dao DC, Osika EM and De Clercq E. Identification of novel thiocarboxanilide derivatives that suppress a variety of drug-resistant mutant human immunodeficiency virus type 1 strains at a potency similar to that for wild-type virus. *Antimicrobial Agents and Chemotherapy* 1996; **40**(6):1454–1466.

[38] Buckheit RW Jr, Fliakas-Boltz V, Decker WD, Roberson JL, Stup TL, Pyle CA, White EL, McMahon JB, Currens MJ, Boyd MR and Bader JP. Comparative anti-HIV evaluation of diverse HIV-1-specific reverse transcriptase inhibitor-resistant virus isolates demonstrates the existence of distinct phenotypic subgroups. *Antiviral Research* 1995; **26**:117–132.

[39] Moeremans M, De Raeymaeker M, Van den Broeck R, Stoffels P, De Brabander M, De Cree J, Hertogs K, Pauwels R, Staszewski S and Andries K. Virological analysis of HIV-1 isolates in patients treated with the non-nucleoside reverse transcriptase inhibitor RO91767, 8-chloro-TIBO. Fourth International Workshop on HIV Drug Resistance, Sardinia, Italy, 6–9 July, 1995. Abstract 33.

[40] Demeter L, Resnick L, Nawaz T, Timpone JG Jr, Batts D and Reichman RC. Phenotypic and genotypic analysis of atevirdine (ATV) susceptibility of HIV-1 isolates obtained from patients receiving ATV monotherapy in a phase I clinical trial (ACTG 187): comparison to patients receiving combination therapy with ATV and zidovudine. Third Workshop on Viral Resistance, 1993, Gaithersburg, MD, USA.

[41] Fan N, Rank KB, Evans DB, Thomas RC, Tarpley WG and Sharma SK. Simultaneous mutations at Tyr-181 and Tyr-188 in HIV-1 reverse transcriptase prevents inhibition of RNA-dependent DNA polymerase activity by the bisheteroarylpiperazine (BHAP) U-90152s. *FEBS Letters* 1995; 370(1–2):59–62.

[42] Balzarini J, Brouwer WG, Felauer EE, De Clercq E and Karlsson A. Activity of various thiocarboxanilide derivatives against wild-type and several mutant human immunodeficiency virus type 1 strains. *Antiviral Research* 1995; **27**:219–236.

[43] Zhang H, Vrang L, Backbro K, Lindz P, Sahlberg C, Unge T and Oberg B. Inhibition of human immunodeficiency virus type 1 wild-type and mutant reverse transcriptase by the phenyl ethyl thiozolyl thiourea derivatives trovirdine and MSG-127. *Antiviral Research* 1995; **28**:331–342.

[44] Vrang L, Rydergard C, Ahgren C, Engelhardt P, Hogberg M, Johansson NG, Kangasmetsa J, Lind P, Noreen R, Sahlberg C, Zhou XX, Karlsson A, Lopez C, Morin Jr JM, Ternansky RJ, Bell FW, Jordan CL, Kinnick MD, Palkowitz JA, Parrish CA, Pranc P, Vasileff RT, West SJ and Oberg B. Comparative rates of in vitro resistance development of HIV-1 to non-nucleoside analog RT inhibitors. *Antiviral Research* 1993; 20(Supplement 1):77.

[45] Demeter LM, Shafer RW, Para M, Morse G, Freimuth W, Merigan TC and Reichman RC. Delavirdine (DLV) susceptibility of HIV-1 isolates obtained from patients receiving DLV monotherapy (ACTG 260). Fourth International Workshop on HIV Drug Resistance, Sardinia, Italy, 6–9 July, 1995. Abstract 23.

[46] Nunberg JH, Schleif WA, Boots EJ, O'Brien JA, Quintero JC, Hoffman JM, Emini EA and Goldman ME. Viral resistance to human immunodeficiency virus type 1-specific pyridinone reverse transcriptase inhibitors. *Journal of Virology* 1991; **65**(9):4887–4892.

[47] Saag MS, Emini EA, Laskin OL, Douglas J, Lapidus WI, Schleif WA, Whitley RJ, Hildebrand C, Byrnes VW, Kappes JC, Anderson KW, Massari FE and Shaw GM. A short-term clinical evaluation of L-697,661, a non-nucleoside inhibitor of HIV-1 reverse transcriptase L-697,661 Working Group. *New England Journal of Medicine* 1993; **329**(15):1065–1072.

[48] Staszewski S, Miller V, Kober A, Colebunders R, Vandercam B, Delescluse J, Clumeck N, Van Wanzeele F, De Brabander M, De Cree J, Moeremans M, Andries K, Boucher C, Stoffels P and Janssen PAJ. Evaluation of the efficacy and tolerance of RO18893, RO89439 (loviride) and placebo in asymptomatic HIV-1-infected patients. *Antiviral Therapy* 1996; **1**:42–50.

[49] Seki M, Sadakata Y, Yuasa S and Baba M. Isolation and characterization of human immunodeficiency virus type-1 mutants resistant to the non-nucleoside reverse transcriptase inhibitor MKC-442. *Antiviral Chemistry and Chemotherapy* 1995; **6**(2):73–79.

[50] Balzarini J, Karlsson A and De Clercq E. Human immunodeficiency virus type 1 drug-resistance patterns with different 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine derivatives. *Molecular Pharmacology* 1993; **44**(4):694–701.

[51] Kleim JP, Winkler I, Rosner M, Kirsch R, Rubsamen-Waigmann H, Paessens A and Reiss G. Different mutational pathways of the HIV-1 RT gene are defined by alternative experimental protocols applied to generate in vitro resistance of HIV-1 to HBY 097. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3–6 July, 1996. Abstract 15.

[52] Vandamme A-M. Polymerase chain reaction (PCR) as a diagnostic tool in HIV infection. Verhandelingen van de Koninklijke Academie voor Geneeskunde van Belgie 1994; **56**(3):231–265.

[53] Balzarini J, Velazquez S, Sanfelix A, Karlsson A, Perez-Perez MJ, Camarasa MJ and De Clercq E. Human immunodeficiency virus type-1 specific purine analogues show a resistance spectrum that is different from that of the human immunodeficiency virus type-1-specific non-nucleoside analogues. *Molecular Pharmacology* 1993; **43**(1):109–114.

References

[54] Balzarini J, Karlsson A, Vandamme AM, Perez-Perez MJ, Zhang H, Vrang L, Oberg B, Backbro K, Unge T and San-Felix A. Human immunodeficiency virus type 1 (HIV-1) strains selected for resistance against the HIV-1-specific [2′,5'-bis-O-(tert-butyldimethylsilyl)-3′-spiro-5'-(4'-amino-1',2′-oxathiole-2′,2′-dioxide)]-β-D-pentofuranosyl (TSAO) nucleoside analogues retain sensitivity to HIV-1-specific nonnucleoside inhibitors. *Proceedings of the National Academy of Sciences, USA* 1993; **90**(15):6952–6956.

[55] Vandamme A-M, Schmit JC, Balzarini J, Van Laethem K, Witvrouw M, Hermans P, Sprecher S, Martinez-Picado J, Clotet B, Peetermans W, Desmyter J and De Clercq E. Presence of TSAO-resistant virus strains in non-experienced patients. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3–6 July, 1996. Abstract 47.

[56] Balzarini J, Pelemans H, Perez-Perez MJ, San-Felix A, Camarasa MJ, De Clercq E and Karlsson A. Marked inhibitory activity of non-nucleoside reverse transcriptase inhibitors against human immunodeficiency virus type 1 when combined with (-)2′,3′-dideoxy-3′-thiacytidine. *Molecular Pharmacology* 1996; 49; (in press).

[57] Balzarini J, Jonckheere H, Harrison WA, Dao DC, Anne J, De Clercq E and Karlsson A. Oxathiin carboxanilide derivatives: a class of non-nucleoside HIV-1-specific reverse transcriptase inhibitors (NNRTIs) that are active against mutant HIV-1 strains resistant to other NNRTIs. *Antiviral Chemistry and Chemotherapy* 1995; **6**:169–178.

[58] Vandamme A-M, Debyser Z, Pauwels R, De Vreese K, Goubau P, Youle M, Gazzard B, Stoffels PA, Cauwenbergh GF, Anne J, Andries K, Janssen PAJ, Desmyter J and De Clercq E. Characterization of HIV-1 strains isolated from patients treated with TIBO R82913. *AIDS Research and Human Retroviruses* 1994; **10**(1):39–46.

[59] de Bethune M-P, Pauwels R, Andries K, Vandamme AM, Peeters M, Colebunders R, Stoffels P, De Clercq E and Desmyter J. AZT resistance reversal by the non-nucleoside reverse transcriptase inhibitor α-APA R18893 in a symptomatic HIV-infected individual. Second HIV Drug Resistance Workshop, Noordwijk, The Netherlands, 3–5 June, 1993. Abstract.

[60] Maass G, Immendoerfer U, Koenig B, Leser U, Mueller B, Goody R and Pfaff E. Viral resistance to the thiazolo-iso-indolinones, a new class of nonnucleoside inhibitors of human immunodeficiency virus type 1 reverse transcriptase. *Antimicrobial Agents and Chemotherapy* 1993; **37**(12):2612–2617.

[61] Nguyen MH, Schinazi RF, Shi C, Goudgaon NM, McKenna PM and Mellors JW. Resistance of human immunodeficiency virus type 1 to acyclic 6-phenylselenenyl- and 6-phenylthiopyrimidines. *Antimicrobial Agents and Chemotherapy* 1994; **38**(10):2409–2414.

[62] Staszewski S, Miller V, Rehmet S, Stark T, De Cree J, De Brabander M, Peeters M, Andries K, Moeremans M, De Raeymaeker M, Pearce G, Van Den Broeck RM, Verbiest W and Stoffels P. Virological and immunological analysis of a triple combination pilot study with loviride, lamivudine and zidovudine in HIV-1-infected patients. *AIDS* 1996; 10(5):F1–F7.

[63] Richman D, Shih CK, Lowy I, Rose J, Prodanovich P, Goff S and Griffin J. Human immunodeficiency virus type 1 mutants resistant to nonnucleoside inhibitors of reverse transcriptase arise in tissue culture. *Proceedings of the National Academy of Sciences, USA* 1991; **88**(24):11241–11245.

[64] Mellors JW, Dutchman GE, Im GJ, Tramontano E, Winkler SR and Cheng YC. In vitro selection and molecular characterization of human immunodeficiency virus-1 resistant to non-nucleoside inhibitors of reverse transcriptase. *Molecular Pharmacology* 1992; **41**(3):446–451.

[65] Buckheit RW Jr, Fliakas-Boltz V, Yeagy-Bargo S, Weislow O, Mayers DL, Boyer PL, Hughes SH, Pan BC, Chu SH and Bader JP. Resistance to 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine derivatives is generated by mutations at multiple sites in the HIV-1 reverse transcriptase. *Virology* 1995; **210**(1):186–193.

[66] Kinjerski TL, Pallansch LA and Buckheit RW Jr. Isolation and characterization of HIV-1 isolates resistant to oxathiin carboxanilide derivatives: Evaluation of variables in the selection process. *Antiviral Chemistry and Chemotherapy* 1996; (in press).

[67] Yang SS, Pattabiraman N, Gussio R, Pallansch L, Buckheit RW Jr and Bader JP. Cross-resistance analysis and molecular modelling of non-nucleoside reverse transcriptase inhibitors targeting drug-resistance Leukaemia 1996; (in press).

[68] Balzarini J, Karlsson A, Sardana VV, Emini EA, Camarasa MJ and De Clercq E. Human immunodeficiency virus 1 (HIV-1)-specific reverse transcriptase (RT) inhibitors may suppress the replication of specific drug-resistant (E138K)RT HIV-1 mutants or select for highly resistant (Y181C to C181I)RT HIV-1 mutants. *Proceedings of the National Academy of Sciences, USA* 1994; **91**(14):6599–6603.

[69] Shaw G, Wei X, Johnson V, Taylor M, Decker J, Kilby M, Lifson J, Hahn B and Saag M. Nucleotide sequence analysis of HIV-1 RNA and DNA from plasma and PBMCs of patients treated with ZDV, ddI and nevirapine: rapid turnover and resistance development in vivo. Third International Workshop on HIV Drug Resistance, Kauai, Hawaii, USA, 2–5 August, 1994. Abstract 71.

[70] Schinazi RF, Lloyd RM Jr, Nguyen M-H, Cannon DL, McMillan A, Ilksoy N, Chu CK, Liotta DC, Bazmi HZ and Mellors JW. Characterization of human immunodeficiency viruses resistant to oxathiolane-cytosine nucleosides. *Antimicrobial Agents and Chemotherapy* 1993; **37**(4):875–881.

[71] Tisdale M, Kemp SD, Parry NR and Larder BA. Rapid in vitro selection of human immunodeficiency virus type 1 resistant to 3′-thiacytidine inhibitors due to a mutation in the YMDD region of reverse transcriptase. *Proceedings of the National Academy of Sciences, USA* 1993; **90**:5653–5656.

[72] Gao Q, Gu Z, Parniak MA, Cameron J, Cammack N, Boucher C and Wainberg MA. The same mutation that encodes low-level human immunodeficiency virus type 1 resistance to 2′,3′-dideoxyinosine and 2′,3′-dideoxycytidine confers high-level resistance to the (-) enantiomer of 2′,3′-dideoxy-3′-thiacytidine. *Antimicrobial Agents and Chemotherapy* 1993; **37**(6):1390–1392.

[73] Larder BA, Kemp SD and Harrigan PR. Potential mechanism for sustained antiretroviral efficacy of AZT-3TC combination therapy. *Science* 1995; **269**:696–699.

[74] Keulen W, van Wijk A, Boucher C and Berkhout B. Initial appearance of 184Ile variant in 3TC-treated patients can be explained by the mutation bias of the HIV-1 RT enzyme. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3–6 July, 1996. Abstract 95.

[75] Gu Z, Gao H, Li X, Parniak MA and Wainberg MA. Novel mutation in the human immunodeficiency virus type 1 reverse transcriptase gene that encodes cross-resistance to 2′,3′-dideoxyinosine and 2′,3′-dideoxycytidine. *Journal of Virology* 1992; **66**(12):7128–7135.

[76] Schinazi RF, Lloyd RM Jr, McMillan A, Gosselin G, Imbach JL and Sommadossi J-P. Development of HIV-1 and SIV resistant to β-L-2′,3′-dideoxycytidine analogues. Fourth International Workshop on HIV Drug Resistance, Sardinia, Italy, 6–9 July, 1995. Abstract 10.

[77] Moeremans M, De Raeymaeker M, Van den Broeck R, Stoffels P and Andries K. Genotypic analysis of HIV-1 isolates from patients receiving loviride alone or in combination with nucleoside reverse transcriptase inhibitor. Fourth International Workshop on HIV Drug Resistance, Sardinia, Italy, 6–9 July, 1995. Abstract 34.

[78] Sharma SK, Fan N, Bank KB, Kopta LA, Olmsted RA, Poppe SM, Slade DE, Thomas RC and Tarpley WG. A drug resistance mutation (G190E) in the inhibitor binding pocket impairs both RNA-dependent DNA polymerase and ribonuclease H activities of HIV-1 reverse transcriptase. Fourth International Workshop on HIV Drug Resistance, Sardinia, Italy, 6–9 July, 1995. Abstract 2.

[79] Kleim JP, Bender R, Kirsch R, Meichsner C, Paessens A, Rosner M, Rubsamen-Waigmann H, Kaiser R, Wichers M, Schneweis KE, Winkler I and Riess G. Preclinical evaluation of HBY 097, a new nonnucleoside reverse transcriptase inhibitor of human immunodeficiency virus type 1 replication. *Antimicrobial Agents and Chemotherapy* 1995; **39**(10):2253–2257.

[80] Kleim J-P, Bender R, Billhardt UM, Meichsner C, Riess G, Rosner M, Winkler I and Paessens A. Activity of a novel quinoxaline derivative against human immunodeficiency virus type 1 reverse transcriptase and viral replication. *Antimicrobial Agents and Chemotherapy* 1993; **37**(8):1659–1664.

[81] Gurusinghe AD, Land SA, Birch C, McGavin C, Hooker DJ, Tachedjian G, Doherty R and Deacon NJ. Reverse transcriptase mutations in sequential HIV-1 isolates in a patient with AIDS. *Journal of Medical Virology* 1995; **46**(3):238–243.

[82] Harrigan PR, Kingborn I, Bloor S, Kemp SD, Najera I, Kohli A, and Larder BA. Significance of amino acid variation at human immunodeficiency virus type 1 reverse transcriptase residue 210 for zidovudine susceptibility. *Journal of Virology* 1996; (in press).

[83] Slade DE, Vavro CL, Stapelton JT, Swack N and St Clair MH. A cysteine at codon 215 of HIV RT confers resistance to ddC. Second HIV Drug Resistance Workshop, Noordwijk, The Netherlands, 3–5 June, 1993. Abstract.

[84] Dueweke TJ, Pushkarskaya T, Poppe SM, Swaney SM, Zhao Q, Chen SY, Stevenson M and Tarpley WG. A mutation in reverse transcriptase of bis(heteroaryl)piperazine-resistant human immunodeficiency virus type 1 that confers increased sensitivity to other nonnucleoside inhibitors. *Proceedings of the National Academy of Sciences, USA* 1993; **90**(10):4713–4717.

[85] Ho DD, Toyoshima T, Mo H, Kempf DJ, Norbeck D, Chen CM, Wideburg NE, Burt SK, Erickson JW and Singh MK. Characterization of human immunodeficiency virus type 1 variants with increased resistance to a C2-symmetric protease inhibitor. *Journal of Virology* 1994; **68**(3):2016–2020.

[86] Tisdale M, Myers R, Parry NR, Oliver N, Machera B and Blair E. Comprehensive analysis of HIV-1 variants individually selected for resistance to six HIV protease inhibitors. Third International Workshop on HIV Drug Resistance, Kauai, Hawaii, USA, 2–5 August, 1994. Abstract 14.

[87] Kaplan AH, Michael SF, Wehbie RS, Knigge MF, Paul DA, Everitt L, Kempf DJ, Norbeck DW, Erickson JW and Swanstrom R. Selection of multiple human immunodeficiency virus type 1 variants that encode viral proteases with decreased sensitivity to an inhibitor of the viral protease. *Proceedings of the National Academy of Sciences, USA* 1994; **91**:5597–5601.

[88] Otto MJ, Reid CD, King RW, Garber S, Baker DB, Anton E and Winslow DL. Exposure of chronically infected PBMCs to DMP 450 can completely suppress virus replication or select resistant variants depending upon the dose of compound. Second National Conference on Human Retroviruses and Related Infections, Washington, DC, USA, 29 January–2 February, 1995. Abstract 464.

[89] Winslow DL, Garber S, Reid C, Anton E and Otto MJ. DMP 450, a new cyclic urea inhibitor of HIV protease with potent in vitro antiviral activity. Eighth International Conference on Antiviral Research, Santa Fe, NM, USA, 23–28 April, 1995. Abstract 22.

[90] Partaledis JA, Yamaguchi K and Byrn RA. In vitro selection and characterization of HIV-1 viral isolates with reduced sensitivity to inhibitors of HIV protease. Third International Workshop on HIV Drug Resistance, Kauai, Hawaii, USA, 2–5 August, 1994. Abstract 8.

[91] Tisdale M, Myers R, Najera I, Kohli A, Kemp S and Larder BA. Analysis of resistance interactions with 141W94 (VX-478) and other HIV-1 protease inhibitors. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3–6 July, 1996. Abstract 27.

[92] King RW, Garber S, Winslow DL, Reid C, Bacheler LT, Anton E and Otto MJ. Multiple mutations in the human immunodeficiency virus protease gene are responsible for decreased susceptibility to protease inhibitors. *Antiviral Chemistry and Chemotherapy* 1995; **6**(2):80–88.

[93] Potts KE, Smidt ML, Stallings WC, Clare M, Pillay D, Richman DD and Bryant ML. In vitro selection and characterization of human immunodeficiency virus type 1 (HIV-1) variants with decreased sensitivity to hydroxyethylurea isostere containing protease inhibitors. Third International Workshop on HIV Drug Resistance, Kauai, Hawaii, USA, 2–5 August, 1994. Abstract 4.

[94] Pillay D, Smidt ML, Potts KE, Bryant ML and Richman DD. In vitro selection of protease inhibitors resistant human immunodeficiency virus type 1 (HIV-1) strains. 34th Interscience Conference on Antimicrobial Agents and Chemotherapy, Orlando, FL, USA, 2–5 October, 1996. Abstract 7.

[95] Condra JH, Holder DJ, Schleif WA, Blahy OM, Danovich RM, Gabryelski LJ, Graham DJ, Laird D, Quintero JC, Rhodes A, Robbins HL, Roth E, Shivaprakash M, Yang T, Chodakewitz JA, Deutsch PJ, Leavitt RY, Massari FE, Mellors JW, Squires KE, SteigbigelRT, Teppler H and Emini EA. Genetic correlates of in vivo viral resistance to the HIV-1 protease indinavir. *Journal of Virology* 1996; (in press).

[96] Schapiro JM, Winters MA, Vierra M, Jacobsen H, Mous J and Merigan TC. Resistance mutations in patients receiving saquinavir: simultaneous appearance in lymph nodes, peripheral blood mononuclears (PBM) and plasma. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3–6 July, 1996. Abstract 28.

[97] Kempf D, Markowitz M, Marsh K, Denissen J, Mo H, Bhat T, Park C, Kong XP, Stewart K, McDonald E, Vasavanonda S, Flentge C, Wideburg N, Robins T, Hsu A, Leonard J, Ho D and Norbeck D. Pharmacokinetic and in vitro selection studies with ABT-538, a potent inhibitor of HIV protease with high oral bioavailability. 34th Interscience Conference on Antimicrobial Agents and Chemotherapy, Orlando, FL, USA, 2–5 October, 1996. Abstract.

[98] Molla A, Korneyeva M, Gao Q, Vasavanonda S, Schipper PJ, Mo HM, Markowitz M, Chernyavskiy T, Niu P, Lyons N, Hsu A, Granneman R, Ho DD, Boucher CAB, Leonard JM, Norbeck DW and Kempf DJ. Ordered accumulation for mutation in HIV protease confers resistance to ritonavir. *Nature Medicine* 1996; **2**(7):760–766.

[99] Doyon I, Croteau G, Thibeault D, Poulin F, Pilocle L and Lamarre D. Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *Journal of Virology* 1996; **70**:3763–3769.

[100] Patick AK, Mo H, Markowitz M, Appelt K, Wu B, Musick L, Kalish V, Kaldor S, Reich S, Ho D and Webber S. Antiviral and resistance studies of AG1343, an orally bioavailable inhibitor of human immunodeficiency virus protease. *Antimicrobial Agents and Chemotherapy* 1996; **40**(2):292–297; **40**(6):1575 (erratum).

[101] Patick AK, Duran M, Cao Y, Ho T, Pei A, Keller MR, Peterkin J, Chapman S, Anderson B, Ho D and Markowitz M. Genotypic and phenotypic characterization of HIV-1 variants isolated from in vitro selection studies and from patients treated with the protease inhibitor, nelfinavir. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3–6 July, 1996. Abstract 29.

[102] Maschera B, Blance C, Brown D and Blair ED. Mutations conferring resistance to HIV-1 protease inhibitors can lie outside, as well as within, the enzyme active site and binding pockets. Third International Workshop on HIV Drug Resistance, Kauai, Hawaii, USA, 2–5 August, 1994. Abstract 9.

[103] Borman AM, Paulous S and Clavel F. Continued accumulation of protease inhibitor resistance mutations in culture in the absence of the drug. Fourth International Workshop on HIV Drug Resistance, Sardinia, Italy, 6–9 July, 1995. Abstract 93.

[104] Lamarre D, Croteau G, Pilote L, Rousseau P and Doyon L. Molecular characterization of HIV-1 variants resistant to specific viral protease inhibitors. Third International Workshop on HIV Drug Resistance, Kauai, Hawaii, USA, 2–5 August, 1994. Abstract 10.

[105] Lamarre D, Doyon L, Croteau G, Pilote L and Thibeault D. Molecular basis of HIV-1 resistance to protease inhibitors Structural flexibility of the protease and second-site compensatory mutations in cleavage sites. Fourth International Workshop on HIV Drug Resistance, Sardinia, Italy, 6–9 July, 1995. Abstract 62.

[106] Gulnik SV, Suvorov LI, Liu B, Yu B, Anderson B, Mitsuya H and Erickson JW. Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry* 1995; **34**(29):9282–9287.

[107] Condra JH and Schleif WA. In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature* 1995; **374**:569–571.

[108] Rao BG, Dwyer MD, Thomson JA, Baker CT, Deininger DD, Murcko MA, Tung RD, Navia MA and Kim EE. Structural and modelling analysis of the basis of viral resistance to VX-478. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3–6 July, 1996. Abstract 22.

[109] Jacobsen H, Brun-Vezinet F, Duncan I, Hanggi M, Ott M, Vella S, Weber J and Mous J. Genotypic characterization of HIV-1 from patients after prolonged treatment with proteinase inhibitor saquinivir. Third International Workshop on HIV Drug Resistance, Kauai, Hawaii, USA, 2–5 August, 1994. Abstract 16.

[110] Eberle J, Bechowsky B, Rose D, Hauser U, Von Der Helm K, Gurtler L and Nitschki H. Resistance of HIV type 1 to proteinase inhibitor Ro 31–8959. *AIDS Research and Human Retroviruses* 1995; **11**(6):671–676.

[111] Swanstrom R, Smith T, Petit S, Irlbeck D, Shao W, Wehbie R, Sawhney R, Everitt L and Erickson I. Multiple sequence changes within HIV-1 protease confer reduced sensitivity to a symmetric protease inhibitor. Third International Workshop on HIV Drug Resistance, Kauai, Hawaii, USA, 2–5 August, 1994. Abstract 6.

[112] Patick AK, Rose R, Greytok J, Bechtol CM, Hermsmeier MA, Chen PT, Barrish JC, Zahler R, Colonno RJ and Lin PF. Characterization of a human immunodeficiency virus type 1 variant with reduced sensitivity to an aminodiol protease inhibitor. *Journal of Virology* 1995; **69**(4):2148–2152.

[113] Rose B, Greytok J, Bechtold C, Alam M, Terry B, Gong YF, DeVore K, Patrick A, Colono R and Lin PF. Combination therapy with two protease inhibitors as an approach to antiviral therapy. Third International Workshop on HIV Drug Resistance, Kauai, Hawaii, USA, 2–5 August, 1994. Abstract 17.

[114] Eastman PS, Gee C, Dewar R, Fyfe G, Metcalf J, Kolberg J, Urdea M, Lane HC and Falloon J. An evaluation of crixivan monotherapy and crixivan/IL-2 combination therapy subjects. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3–6 July, 1996. Abstract 34.

[115] Otto MJ, Garber S, Winslow DL, Reid CD, Aldrich P, Jadhav PK, Patterson CE, Hodge CN and Cheng YS. In vitro isolation and identification of human immunodeficiency virus (HIV) variants with reduced sensitivity to C-2 symmetrical inhibitors of HIV type 1 protease. *Proceedings of the National Academy of Sciences, USA* 1993; **90**(16):7543–7.

[116] Shao W, Smith T and Swanstrom R. Selection and analysis of HIV-1 variants with increased resistance to SKF108842 and SKF108922, two protease inhibitors. Fourth International Workshop on HIV Drug Resistance, Sardinia, Italy, 6–9 July, 1995. Abstract 65.

[117] el-Farrash MA, Kuroda MJ, Kitazaki T, Masuda T, Kato K, Hatanaka M and Harada S. Generation and characterization of a human immunodeficiency virus type 1 (HIV-1) mutant resistant to an HIV-1 protease inhibitor. *Journal of Virology* 1994; **68**(1):233–9.

[118] Este JA, Van Laethem K, Vandamme AM, Desmyter J and De Clercq E. Resistant phenotype of human immunodeficiency virus type 1 to dextran sulfate is conferred by specific amino acid substitutions in the gp120 molecule. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3–6 July, 1996. Abstract 80.

[119] Este JA, Schols D, De Vreese K, Van Laethem K, Vandamme AM, Desmyter J and De Clercq E. Development of resistance of human immunodeficiency virus type 1 to dextran sulfate associated with the emergence of specific mutations in the envelope gp120 glycoprotein. Submitted for publication 1996.

[120] De Vreese K, Reymen D, Griffin P, Steinkasserer A, Werner G, Bridger GJ, Este J, James W, Henson GW, Desmyter J, Anne J and De Clercq E. The bicyclams, a new class of potent human immunodeficiency virus inhibitors, block viral entry after binding. *Antiviral Research* 1996; **29**:209–219.

[121] De Vreese K, Kofler-Mongold V, Leutgeb C, Weber V, Vermeire K, Schacht S, Anne J, De Clercq E, Datema R and Werner G. The molecular target of bicyclams, potent inhibitors of human immunodeficiency virus replication. *Journal of Virology* 1996; **70**(2):689–696.

[122] Lin P, Samanta H, Bechtold CM, Deminie CA, Patick AK, Alam M, Riccardi K, Rose RE, White RJ and Colonno RJ. Characterization of siamycin 1, a human immunodeficiency virus fusion inhibitor. *Antimicrobial Agents and Chemotherapy* 1995; **40**:133–138.

[123] Iversen AK, Shafer RW, Wehrly K, Winters MA, Mullins JI, Chesebro B and Merigan TC. Multidrug-resistant human immunodeficiency virus type 1 strains resulting from combination antiretroviral therapy. *Journal of Virology* 1996; **70**(2):1086–1090.

[124] Shirasaka T, Kavlick MF, Ueno T, Gao WY, Kojima E, Alcaide ML, Chokekijchai S, Roy BM, Arnold E, Yarchoan R and Mitsuya H. Emergence of human immunodeficiency virus type 1 variants with resistance to multiple dideoxynucleosides in patients receiving therapy with dideoxynucleosides. *Proceedings of the National Academy of Sciences, USA* 1995; **92**:1–5.

[125] Schmit JC, Vanderlinden I, Ruiz L, Clotet B, Hermans P, Sprecher S, Arendt V, Peetermans W, Harrer T, Vaira D, Desmyter J, De Clercq E and Vandamme AM. Prevalence of multi-drug resistance to dideoxynucleoside (ddN) analogues in patients on ddN combination therapy. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3–6 July, 1996. Abstract 39.

IV

Related Sequences

HIV and Cellular Factors

Kuan-Teh Jeang

Molecular Virology Section, LMM, NIAID, National Institutes of Health, Bethesda, MD 20892-0460, USA

Introduction

Viruses are obligatory parasites of cells. Thus it is expected that host cell factors contribute importantly to the life-cycle of viruses. In this section, we survey, for HIV-1, four types of virus-cell interactions. These four areas include: 1) DNA-binding proteins that recognize target motifs in proviral LTR; 2) RNA-binding proteins that bind HIV-1 RNAs; 3) cellular factors that form protein-protein complexes with HIV-1 regulatory proteins; and 4) cellular genes which are modulated upon viral infection. With rapid increases in knowledge in the area of virus-cell interactions, we anticipate that this initial survey would be expanded extensively in future editions of the data base. Additional discussions on HIV-cell interactions are found elsewhere (Jones and Peterlin, 1994; Jeang and Gatignol, 1994; Garcia and Gaynor, 1994; Gatignol *et al.*, 1996; Dayton, 1996).

I. DNA-binding Proteins

The promoter-enhancers of the human immunodeficiency virus (HIV) are contained in the U3 of the viral long terminal repeat (LTR). HIV-1 U3 is typically 454 nucleotides long and has binding sites for many transcription factors. Some of these are diagrammed in figure 1. With the exception of NF-kB and Sp1 (Ross *et al.*, 1991; Kim *et al.*, 1993; Huang and Jeang, 1993; Huang *et al.*, 1994), verification of the importance of the various sequences in viral contexts has not been directly performed.

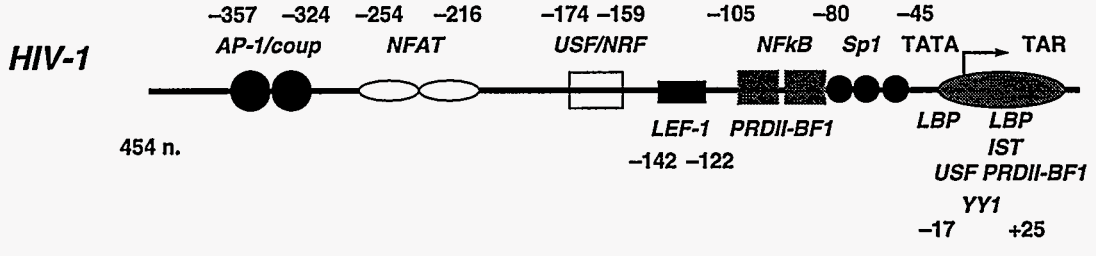


Fig. 1. Diagrammatic representation of the positions in U3 and R for some of the DNA-binding proteins that recognize the HIV-1 LTR.

Subgenomic assays in cultured cells indicate that the primary contributors to HIV-promoter activity are the NF-kB, Sp1, and TATAA (sequences from +1 to –105; Berkhout and Jeang, 1992). Nonetheless, it is likely that other DNA-binding factors also contribute *in vivo*. Beginning directionally from the 5′ end of U3, binding sites for AP-1 (a fos/jun hetero-complex; van Straaten *et al.*, 1983; Hattori *et al.*, 1988) and COUP (a member of the steroid/thyroid receptor superfamily; Cooney *et al.*, 1991) are found between –324 to –357. Both COUP and AP-1 are expressed in human T-cells, and thus could compete with each other for the same proviral DNA sites during HIV-1 infection.

More proximally, between –216 and –254 are binding motifs for nuclear factor from activated T-cells (NFAT; Shaw *et al.*, 1988). NFAT is an intermediating transducer of signals initiated at the T-cell antigen receptor. Recent evidence suggests that NFAT binding activity is composed of three discrete polypeptides, NFATp (McCaffrey *et al.*, 1993), Fos and Jun (Yaseen *et al.*, 1993).

In the region between –159 to –174 is a binding consensus sequence for USF (Gregor *et al.*, 1990). USF was characterized initially as a positive activator of adenovirus major-late-promoter transcription. In the HIV context, there is conflicting information on whether this factor has moderating (and thus be regarded as a negative regulatory factor; NRF; Lu *et al.*, 1990) or stimulating (Maekewa *et al.*, 1991) effects. Interestingly, USF also binds a second unrelated sequence (–5 to +11) that surrounds the HIV-1 initiator (Hu *et al.*, 1993). USF interaction at the initiator-proximal site activates strong expression from the TATAA-promoter (Hu *et al.*, 1993). Of note, a factor distinct from USF, but which binds the same DNA-sequence, has also been cloned and characterized (TFE3; Bechmann and Kadesch, 1991).

LEF-1 is a T-cell specific transcription factor (Waterman and Jones, 1990). Once bound to its cognate site, LEF-1 bends DNA and thereby facilitates the assembly of nucleoprotein complexes at the promoter (Giese *et al.*, 1992; reviewed in Jones and Peterlin, 1994). A high affinitiy LEF-1 binding site is present at –122 to –143. Two low affinity binding sites exist at –37 to –51 and +17 to +32 (Waterman and Jones, 1990).

NF-kB (Nabel and Baltimore, 1987) and Sp1 (Jones and Tjian, 1985; Jones *et al.*, 1986) motifs are perhaps the best characterized sequence elements in the HIV-1 LTR. These sequences directly impact viral replication (Ross et al, 1991; Kim *et al.*, 1993; Huang and Jeang, 1993; Huang *et al.*, 1994), viral transcription (Harrich *et al.*, 1990; Berkhout and Jeang, 1992; Pazin *et al.*, 1996), and Tat transactivation (reviewed in Jones and Peterlin, 1994; Jeang and Gatignol, 1994). For more extensive discussions of the biochemical and functional properties of NF-kB (Ghosh *et al.*, 1990; Kieran *et al.*, 1990; Nolan *et al.*, 1991; Liou *et al.*, 1991; and references cited therein) and Sp1 (Dynan and Tjian, 1983; Briggs *et al.*, 1986; Kadonaga *et al.*, 1987; and references cited therein), readers should consult elsewhere.

PRDII-BF1 is a 300 kDa zinc-finger containing protein (Baldwin *et al.*, 1990; Seeler *et al.*, 1994). PRDII-BF1 recognizes and binds the NF-kB motif; however, it also binds a divergent sequence in R (+27 to +52; Seeler *et al.*, 1994).

Positioned at the junction of U3 and R are sites for LBP (Yoon *et al.*, 1994) and YY1 (Useheva and Shenk, 1994; Seto *et al.*, 1991). While the role for LBP in HIV-1 transcription is not wholly understood (Jones *et al.*, 1988; Kato *et al.*, 1991), binding of YY1 to the LTR has been shown to repress HIV-1 expression and production of virions (Margolis *et al.*, 1994). Besides LBP and YY1, a DNA-mediated activity for the induction of short transcripts (IST) has also been mapped to the same general vicinity (–5 to +26; Sheldon *et al.*, 1993). The cDNA for the cellular factor that mediates IST-activity has not been isolated, and thus the authentic identity of this factor is unknown. HIP 116 is another newly cloned cDNA that binds to the TATA/initiator of the HIV-1 promoter (Sheridan *et al.*, 1995).

There are a number of reports of sites for DNA-binding proteins in the HIV-1 genome that occur downstream of the +1 start for transcription. These include NF-kB motifs (Mallardo *et al.*, 1996) and sites for AP-1, AP-3-like, DBF-1, and Sp-1 (El Kharroubi and Martin, 1996). The full implication of these downstream DNA-motifs in the setting of HIV-1 infection remains to be clarified.

II. RNA-binding Proteins

In recent years, it has become evident that RNA-binding proteins play important roles in gene regulation (see reviews, Keene and Query, 1991; Mattaj, 1993; Burd and Dreyfuss, 1994). For HIV-1, cellular proteins that bind viral regulatory RNAs have been studied in detail. In particular, at least eight host cell factors have been described to bind TAR RNA. Similarly, two RRE-binding factors have been characterized. There are also biologically compelling reasons as to why TAR- and RRE- binding proteins are meaningful contributors to the HIV-1 lifecycle.

The HIV-1 leader RNA, TAR, forms a stem-bulge-loop structure of approximately 60+ nucleotides (Muesing *et al.*, 1987; Berkhout and Jeang, 1989). Early, it was reported that many human cellular proteins bound TAR RNA (Gatignol *et al.*, 1989; Gaynor *et al.*, 1989). Since then, some of these factors have been defined further. Tabulating from extant studies, eight proteins associate with either the bulge, loop, or stem of TAR RNA (see fig. 2). TAR loop-binding proteins include p68 (Marciniak *et al.*, 1990), and TRP1/TRP185 (185 kDa; Sheline *et al.*, 1991; Wu *et al.*, 1991), while TRP2 (70–110 kDa; Sheline *et al.*, 1991) binds to TAR-bulge. Proteins that complex with the double-stranded stem of TAR RNA

consist of P1/dsI (newly renamed as PKR; McCormack et al., 1992; Roy *et al.*, 1991), SBP (140 kDa; Rounseville and Kumar, 1992), and TRBP (Gatignol *et al.*, 1991; Gatignol *et al.*, 1993). Two human autoantigens have been identified as TAR RNA-binding factors: Lupus antigen Ku (Kaczmarski and Khan, 1993) binds to the loop of TAR, while La (Chang *et al.*, 1994; Svitkin *et al.*, 1994) recognizes U-residues within the overall context of the TAR secondary structure.

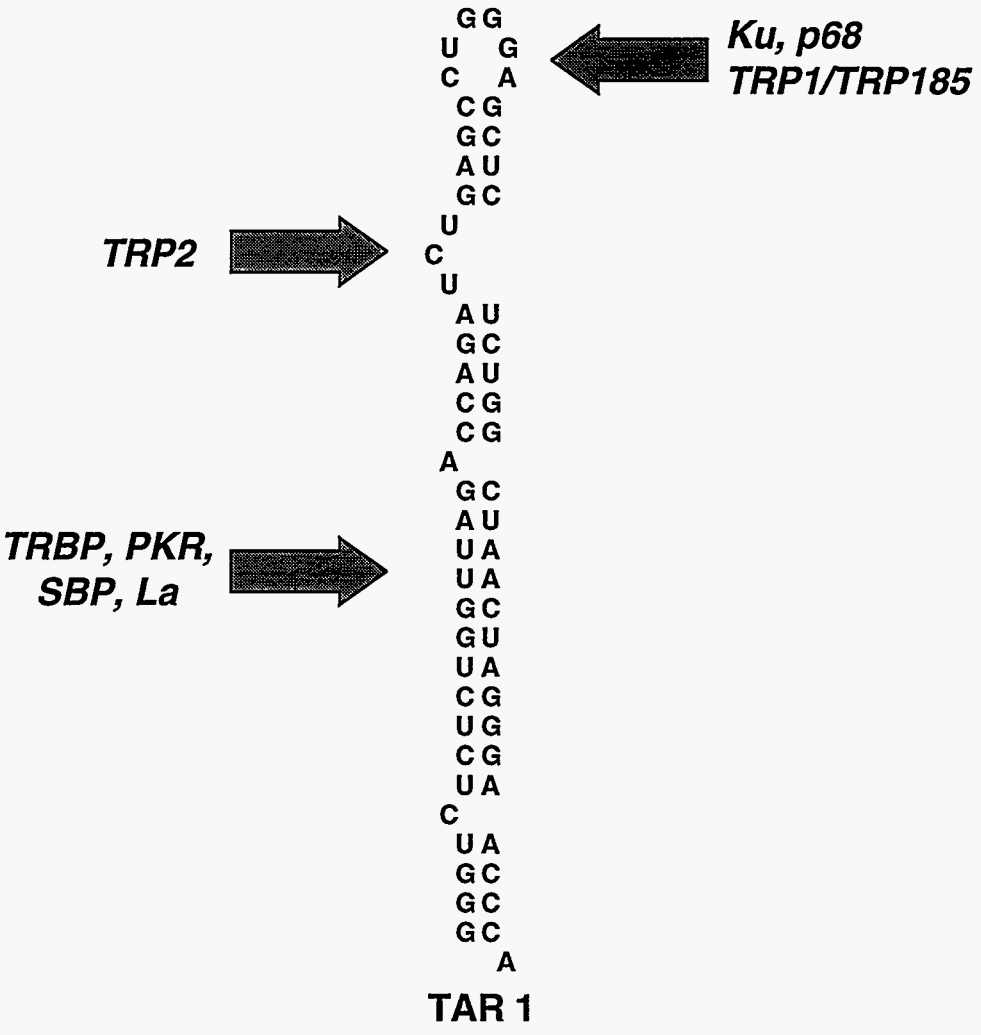


Fig. 2. Structure of the HIV-1 TAR RNA. Identities of the proteins that bind to the loop, bulge or stem of the TAR hairpin are indicated.

Recently, additional TAR-RNA-binding proteins and co-factors have been reported. Wong-Staal and colleagues have described the binding of TARBP-b (Reddy *et al.*, 1995) to the bulge structure of TAR. A set of cellular co-factors that enhances the binding of TRP 185 and RNA polymerase II (RNAP II) to TAR have been described by Gaynor and colleagues (Wu-Baer *et al.*, 1996a; 1996b). Finally, Hart and co-workers have characterized the elusive human chromosome 12-associated TAR loop-binding factor as an 83-kilodalton primate cell-specific protein whose expression is constitutively absent in hamster cells (Hart *et al.*, 1995).

RRE-binding proteins have also been studied extensively. We know that TRBP also can bind RRE (Park *et al.*, 1994). A 56 kDa factor (Vaishnav *et al.*, 1992) and a 49 kDa protein that belongs to

the heterogeneous nuclear ribonuclear protein family (Xu *et al.*, 1996) have been reported to bind RRE. Two other factors of 120 kDa and 60 kDa in molecular size have been reported to be primate-specific RRE-binding factors that are not expressed in rodent cells (Shukla *et al.*, 1994). It is expected that other RRE-factors would emerge from future studies. How these, as yet poorly characterized, factors might contribute to Rev/RRE function is reviewed in detail elsewhere (Dayton, 1996).

III. Protein-protein interactions

Protein-protein interactions are well-documented to be important in gene regulation (see reviews, Lewin, 1990; Greenblatt, 1991). Recent studies suggest that many HIV-1 proteins complex with host cell factors. We briefly describe below some examples pertaining to Tat, Rev, Gag, and Nef.

Although Tat is best known for transcription, it has other functions (Huang *et al.*, 1994; reviewed in Chang *et al.*, 1995; Goldstein, 1996) and has been reported to be a secretable factor that promotes the growth of Kaposi-like cells (Ensoli *et al.*, 1990). There is evidence that Tat can be taken up actively into cells (Frankel and Pabo, 1988) through binding to a cell-surface protein (Weeks *et al.*, 1993) implicated to be $\alpha_v\beta_5$ integrin (Vogel *et al.*, 1993). Once inside cells, Tat interacts with multiple partners in activating transcription (see also Tat Structure and Function section; Part III).

Two factors that bind Tat are themselves critical components of the eucaryotic RNA polymerase II transcription machinery. Genetic evidence supports a critical role for Sp1 in HIV-1 Tat-mediated transactivation (Harrich *et al.*, 1989; Kamine *et al.*, 1991; Southgate and Green, 1991; Berkhout and Jeang, 1992). Interestingly, direct protein-protein interactions between Tat and Sp1 (Jeang *et al.*, 1993) and Tat and the thyroid hormone receptor (which in certain cells bind at the Sp1 sites in the HIV-1 LTR; Desai-Yajnik *et al.*, 1995) have been documented. In addition, there is evidence that Tat also contacts TBP (Kashanchi *et al.*, 1994) and that Tat-Sp1-TFIID could present as a multiprotein complex (Huang *et al.*, 1993) inside cells. The critical nature of Tat-TBP contact has been questioned recently (Wang *et al.*, 1996). These investigators found that a trans-activation defective Tat protein bound TBP with the same avidity as wild type Tat protein, suggesting that binding to TBP cannot solely reflect the trans-activation property of Tat.

A large family of proteins related to the 26S protease from human erythrocytes (Dubiel *et al.*, 1994) are Tat-binding polypeptides. Members of this family include TBP-1 (Nelbock *et al.*, 1990; Ohana *et al.*, 1993), TBP-7 (Ohana *et al.*, 1993; Shaw and Ennis, 1993), MSS1 (Shibuya *et al.*, 1992), and SUG1 (Swaffield *et al.*, 1992). The exact role of this family of proteins in cellular metabolism is not wholly clear; however, the 26S protease seems to regulate the degradation of some cyclins and in this manner has been implicated in modulating the stability of oncoproteins such as c-Mos, c-Myb, c-Myc and p53 (Dubiel *et al.*, 1994).

Other cellular proteins, a 36 kDa protein (Desai *et al.*, 1991), a 60 kDa protein (Kamine *et al.*, 1996), and different cellular protein kinases (Hermann and Rice, 1993; McMillan *et al.*, 1995; Zhou and Sharp, 1996; Chun and Jeang, 1996), have been reported to bind Tat. Mavankal and colleagues have suggested that RNA polymerase II (RNAP II) is itself a specific Tat-binding protein (Mavankal *et al.*, 1996). The findings that RNAP II and kinases that phosphorylate RNAP II are Tat-binding factors suggest that these components contribute a part to explaining Tat function in transcription.

Multiple cellular factors also interact with Rev. Rev localizes to the nucleoli, and work by Laemmli and colleagues (Fankhauser *et al.*, 1991) and others (Miyazaki *et al.* 1995; Szebeni *et al.*, 1995) have demonstrated a tight association between the basic nuclear-localizing domain of Rev and nucleolar B23 protein. This protein-protein complex likely directs the subcellular localization of Rev.

By contrast, many investigators have studied the nuclear export function of Rev (Fischer *et al.*, 1994; 1995; Wen *et al.*, 1995; Myer *et al.*, 1996; Fridell *et al.*, 1996). Using the yeast two-hybrid cloning approach, three groups have isolated closely related nucleoporin proteins that bind wild type but not mutant Rev (Fritz *et al.*, 1995; Bogerd et al., 1995; Stutz *et al.*, 1995). These results suggest that the nucleoporin class of proteins are the Rev activation-domain specific co-factors involved in the nuclear to cytoplasmic export of RNAs.

Rev function can also be modulated by associations with other cellular factors. Two additional proteins have been shown to bind Rev. These are translation factor eIF-5A (Ruhl *et al.*, 1993; Bevec *et al.*, 1996), and serine-arginine (SR)-rich factors (Gontarek and Derse, 1996) such as the p32 protein (Luo *et al.*, 1994; Tange *et al.*, 1996) originally characterized by Krainer and colleagues (Krainer *et al.*, 1991) as a splicing factor-SF2-associated polypeptide. SR-rich proteins can also bind to the basic domain of the HIV-1 and HIV-2 Tat proteins (Trinh and Jeang, unpublished observation; B.R. Cullen, personal communication).

Other less-characterized Rev-binding proteins include human prothymosin alpha (Kubota et al, 1995) and nuclear lamin B (Nikolakaki et al, 1996).

Both HIV Gag and Nef also have cellular partners. p55 and p24 Gag bind cyclophilin A and cyclophilin B proteins (Luban *et al.*, 1993). The cyclophilins are cellular polypeptides originally characterized for their specific binding to cyclosporin A.

Nef has been found to bind β-COP, a coat protein from non-clathrin-coated vesicles (Benichou *et al.*, 1994). This interaction may be important in promoting the intracellular sequestration of CD4. Nef has also been shown to interact directly with CD4 (Rossi *et al.*, 1996). Recently many kinases have been described to associate with Nef (Sawai *et al.*, 1994; Saksela *et al.*, 1995; Bodeus *et al.*, 1995; Luo and Garcia, 1996). The identity of some of these kinases include Lck (Salghetti *et al.*, 1995; Collette *et al.*, 1996), Hck (Lee *et al.*, 1995), and PAK (Nunn and Marsh, 1996). Undoubtedly, there is much more to be learned from this area of research.

IV. Cellular Genes Modulated by HIV

Infection of cells by HIV results in activation and repression of many cellular genes. Because of the complexity of ambient gene expression inside cells, it is logistically difficult to dissect those genes that are upregulated from those that are downregulated from others that remain unperturbed. Nevertheless, several examples of genes that respond to HIV infection are known. Some of these include IL-2 (Westendorf *et al.*, 1994), IL6 (Scala *et al.*, 1994), and TGF-β (Buonaguro *et al.*, 1994; Rasty *et al.*, 1996).

Recent investigations have shown that HIV-1 proteins such as Tat and Nef potently affect cellular metabolism. Tat, for example, has been shown to modulate the expression of cytokines including MIP-1 alpha (Sharma *et al.*, 1996), second messengers including nitric oxide (Barton *et al.*, 1996), and housekeeping genes such as Bcl2 (Zauli *et al.*, 1995a). Both Tat and Nef participate in signal transduction pathways that include NF-kB (Demarchi *et al.*, 1996), phosphatidylinositol kinase (Zauli et al; 1995b; Gramagli et al., 1996), and protein kinase C (Conant *et al.*, 1996)

The identification of genes modulated by HIV has been based on classical approaches; however, with the advent of mRNA differential display technology (Liang and Pardee, 1992) the exhaustive characterization of cellular genes that respond to infection by HIV should be accomplished with greater ease. One such example has recently been reported. Zeichner and colleagues, using a differential display approach, found that expression of the human glucose transporter protein (GLUT3) is upregulated by HIV-1 infection. We anticipate reports on others, shortly.

References

[1] Baldwin, A. S. J., K. P. LeClair, H. Singh, and P. A. Sharp. 1990. A large protein containing zinc finger domains binds to related sequence elements in the enhancers of the class I major histocompatibility complex and kappa immunoglobulin genes. *Mol. Cell. Biol.* **10**:1406–1414.

[2] Barton, C. H., T. E. Biggs, T. R. Mee, and D. A. Mann. 1996. The human immunodeficiency virus type 1 regulatory protein Tat inhibits interferon-induced iNos activity in a murine macrophage cell line. *J. Gen. Virol.* **77**(Aug Pt 8):1643–1647.

[3] Beckmann, H., and T. Kadesch. 1991. The leucine zipper of TEF3 dictates helix-loop-helix dimerization specificity. *Genes Dev.* **5**:1057–1066.

[4] Benichou, S., M. Bomsel, M. Bodeus, H. Durand, M. Doute, F. Letourneur, J. Camonis, and R. Benarous. 1994. HIV-1 nef interacts with B-COP, a coat protein from non-clathrin-coated vesicles. *J. Biol. Chem.* **269**: 30073–30076.

[5] Berkhout, B., and K.-T. Jeang. 1989. trans Activation of human immunodeficiency virus type 1 is sequence specific for both the single-stranded bulge and loop of the trans-acting-responsive hairpin: a quantitative analysis. *J. Virol.* **63**:5501–5504.

[6] Berkhout, B., and K.-T. Jeang. 1992. Functional roles for the TATA promoter and enhancers in basal and Tat-induced expression of the Human Immunodeficiency virus type 1 long terminal repeat. *J. Virol.* **66**:139–149.

[7] Berkhout, B., R. H. Silverman, and K.-T. Jeang. 1989. Tat trans-activates the human immunodeficiency virus through a nascent RNA target. *Cell* **59**:273–282.

[8] Bevec, D., H. Jaksche, M. Oft, T. Wohl, M. Himmelspach, A. Pacher, M. Schebesta, K. Koettnitz, M. Dobrovnik, R. Csonga, F. Lottspeich, and J. Hauber. 1996. Inhibition of HIV-1 replication in lymphocytes by mutants of the Rev cofactor eIF-5A. *Science* **271**:1858–1860.

[9] Bodeus, M., A. Marie-Cardine, C. Bougeret, F. Ramos-Morales, and R. Benarous. 1995. *In vitro* binding and phosphorylation of human immunodeficiency virus type 1 Nef protein by serine/threonine protein kinase. *J. Gen. Virol.* **76**:1337–1344.

[10] Bogerd, H. P., R. A. Fridell, S. Madore, and B. R. Cullen. 1995. Identification of a novel cellular cofactor for the Rev/Rex class of retroviral regulatory proteins. *Cell* **82**:485–494.

[11] Braddock, M., R. Powell, A. D. Blanchard, A. J. Kingsman, and S. M. Kingsman. 1993. HIV-1 TAR RNA-binding proteins control TAT activation of translation in Xenopus oocytes. *FASEB J.* **7**:214–222.

[12] Briggs, M. R., J. T. Kadonaga, S. P. Bell, and R. Tjian. 1986. Purification and biochemical characterization of the promoter specific transcription factor, Sp1. *Science* **234**:47–52.

[13] Buonaguro, L., F. M. Buonaguro, G. Giraldo, and B. Ensoli. 1994. The human immunodeficiency virus type 1 Tat protein transactivates tumor necrosis factor beta gene expression through a TAR-like structure. *J. Virol.* **68**:2677–2682.

[14] Burd, C. G., and G. Dreyfuss. 1994. Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**:615–621.

[15] Chang, H. K., R. C. Gallo, and B. Ensoli. 1995. Regulation of cellular gene expression and function by the human immunodeficiency virus type 1 Tat protein. *J. Biomed. Sci* **2**:189–202.

[16] Chang, Y., D. J. Kenan, J. Keene, A. Gatignol, and K. T. Jeang. 1994. Direct interactions between the autoantigen La and the human immunodeficiency virus (HIV-1) leader RNA. *J. Virol.* **68**:7008–7020.

[17] Chun, R., and K.-T. Jeang. 1996. Requirements for RNA polymeraseII carboxyl-terminal domain for activated transcription of human retroviruses human T-cell lymphotropic virus I and HIV-1. *J. Biol. Chem.* **271**:27888–27894.

[18] Collette, Y., H. Dutartre, A. Benziane, A. Romas-Morales, R. Benarous, M. Harris, and D. Olive. 1996. Physical and functional interaction of Nef with Lck. HIV-1 Nef-induced T-cell signaling defects. *J. Biol. Chem.* **271**:6333–6341.

[19] Conant, K., M. Ma, A. Nath, and E. O. Major. 1996. Extracellular human immunodeficiency virus type 1 Tat protein is associated with an increase in both NF-kappa B binding and protein kinase C activity in primary human astrocytes. *J. Virol.* **70**:1384–1389.

[20] Cooney, A. J., S. Y. Tsai, B. W. O'Malley, and M.-J. Tsai. 1991. Chicken ovalbumin upstream promoter transcription factor binds to a negative regulatory region in the human immunodeficiency virus type 1 long terminal repeat. *J. Virol.* **65**:2853–2860.

[21] Dayton, A. I. 1996. The Rev axis and its associated host cofactors: a viral window onto the workings of eukaryotic posttranscriptional RNA processing. *Journal of Biomedical Science* **3**:69–77.

[22] Demarchi, F., di F. F. d'Adda, A. Falaschi, and M. Giacca. 1996. Activation of transcription factor NF-kappaB by the Tat protein of human immunodeficiency virus type 1. *J. Virol.* **70**:4427–4437.

[23] Desai, K., P. M. Lowenstein, and M. Green. 1991. Isolation of a cellular protein that binds to the human immunodificiency virus Tat protein and can potentiate transactivation of the viral promoter. *Proc. Natl. Acad. Sci. USA* **88**:8875–8879.

[24] Desai-Yajnik, V., E. Hadzic, P. Modlinger, S. Malhotra, G. Gechlik, and H. H. Samuels. 1995. Interactions of thyroid hormone receptor with the human immunodeficiency virus type 1 (HIV-1) long terminal repeat and the HIV-1 Tat transactivator. *J. Virol.* **69**:5103–5112.

[25] Du, H., A. L. Roy, and R. G. Roeder. 1993. Human transcription factor USF stimulates transcription through the initiator elements of the HIV-1 and the Ad-ML promoters. *EMBO J.* **12**:501–511.

[26] Dubiel, W., K. Ferrell, and M. Rechsteiner. 1994. Tat-binding protein 7 is a subunit of the 26S protease. *Biol. Chem. Hoppe-Seyler* **375**:237–240.

[27] Dynan, W. S., and R. Tjian. 1983. Isolation of transcriptionfactors that discriminate between different promoters recognized by RNA polymerase II. *Cell* **32**:669–680.

[28] El Kharroubi, A., and M. A. Martin. 1996. cis-Acting sequences located downstream of the human immunodeficiency virus type 1 promoter affect its chromatin structure and transcriptional activity. *Mol. Cell. Biol.* **16**:2958–2966.

[29] Ensoli, B., G. Barillari, S. Z. Salahuddin, R. C. Gallo, and F. Wong-Staal. 1990. Tat protein of HIV-1 stimulates growth of cells derived from Kaposi's sarcoma lesions of AIDS patients. *Nature* **345**:84–86.

[30] Fankhauser, C., E. Izaurralde, Y. Adachi, P. Wingfield, and U. K. Laemmli. 1991. Specific complex of human immunodeficiency virus type 1 Rev and nucleolar B23 proteins: dissociation by the Rev response element. *Mol. Cell. Biol.* **11**:2567–2575.

[31] Fischer, U., J. Huber, W. C. Boelens, I. W. Mattaj, and R. Luhrmann. 1995. The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell* **82**:475–483.

[32] Fischer, U., S. Meyer, M. Teufel, C. Heckel, R. Luhrmann, and G. Rautmann. Evidence that HIV-1 Rev directly promotes the nuclear export of unspliced RNA. *EMBO J.* **13**:4105–4112.

[33] Frankel, A., and C. Pabo. 1988. Cellular uptake of the Tat protein from human immunodeficiency virus. *Cell* **55**:1189–1193.

[34] Fridell, R. A., H. P. Bogerd, and B. R. Cullen. 1996. Nuclear export of late HIV-1 mRNAs occurs via a cellular protein export pathway. *Proc. Natl. Acad. Sci. USA* **93**:4421–4424.

[35] Fritz, C. C., M. L. Zapp, and M. R. Green. 1995. A human nucleoporin-like protein that specifically interacts with HIV Rev. *Nature* **376**:530–533.

[36] Garcia, J. A., and R. B. Gaynor. 1994. Regulatory mechanism involved in the control of HIV-1 gene expression. *J. AIDS* 8:S3–S17.

[37] Garcia, J. A., D. Harrich, E. Soultanakis, F. Wu, R. Mitsuyasu, and R. B. Gaynor. 1989. Human immunodeficiency virus type 1 LTR TATA and TAR region sequences required for transcriptional regulation. *EMBO J.* **8**:765–778.

[38] Garcia, J. A., F. K. Wu, R. Mitsuyasu, and R. B. Gaynor. 1987. Interactions of cellular proteins involved in the transcriptional regulation of the human immunodeficiency virus. *EMBO J.* **6**:3761–3770.

[39] Gatignol, A., C. Buckler, and K.-T. Jeang. 1993. Relatedness of an RNA binding motif in HIV-1 TAR RNA-binding protein TRBP to human P1/dsI kinase and Drosophila staufen. *Mol. Cell. Biol.* **13**:2193–2202.

[40] Gatignol, A., A. Buckler-White, B. Berkhout, and K.-T. Jeang. 1991. Characterization of a human TAR RNA-binding protein that activates the HIV-1 LTR. *Science* **251**:1597–1600.

[41] Gatignol, A., M. Duarte, L. Daviet, Y. N. Chang, and K. T. Jeang. 1996. Sequential steps in Tat trans-activation of HIV-1 mediated through cellular DNA, RNA, and protein binding factors. *Gene Exp.* **5**:217–228.

[42] Gatignol, A., A. Kumar, A. Rabson, and K.-T. Jeang. 1989. Identification of cellular proteins that bind to the human immunodeficiency virus type 1 trans-activation-responsive TAR element RNA. *Proc. Natl. Acad. Sci. USA* **86**:7828–7832.

[43] Gaynor, R., E. Soultanakis, M. Kuwabara, J. Garcia, and D. S. Sigman. 1989. Specific binding of a HeLa cell nuclear protein to RNA sequences in the human immunodeficiency virus transactivating region. *Proc. Natl. Acad. Sci. USA* **86**:4858–4862.

[44] Ghosh, S., A. M. Gifford, L. R. Fiviere, P. Tempst, G. P. Nolan, and D. Baltimore. 1990. Cloning of the p50 DNAbinding subunit of NF-kB: homology to rel and dorsal. *Cell* **62**:1019–1029.

[45] Giese, K., J. Cox, and R. Gosschedl. 1992. The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell* **69**:185–195.

[46] Goldstein, G. 1996. HIV-1 Tat protein as a potential AIDS vaccine. *Nat. Med.* **1**:960–964.

[47] Gontarek, R. R., and D. Derse. 1996. Interactions among SR proteins, an exonic splicing enhancer, and a lentivirus Rev protein regulate alternative splicing. *Mol. Cell. Biol.* **16**:2325–2331.

[48] Goyer, C., H. S. Lee, D. Malo, and N. Sonenberg. 1992. Isolation of a yeast gene encoding a protein homologous to the human Tat-binding protein, TBP-1. *DNA Cell Biol.* **13**:579–585.

[49] Graziani, A., F. Galimi, E. Medico, E. Cottone, D. Gramaglia, C. Boccaccio, and P. M. Comoglio. 1996. The HIV-1 nef protein interferes with phosphatidylinositol 3-kinase activation 1. *J. Biol. Chem.* **271**:6590–6593.

[50] Greenblatt, J. 1991. Roles of TFIID in transcriptional initiation by RNA polymerase II. *Cell* **66**:1067–1070.

[51] Gregor, P. D., M. Sawadago, and R. G. Roeder. 1990. The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. *Genes Dev.* **4**:1730–1740.

[52] Han, X. M., A. Laras, M. P. Rouseville, A. Kumar, and P. R. Shank. 1992. Human immunodeficiency virus type 1 Tat-mediated trans activation correlates with the phosphorylation state of a cellular TAR RNA stem-binding factor. *J. Virol.* **66**:4065–4072.

[53] Harrich, D., J. Garcia, F. Wu, R. Mitsuyasu, J. Gonzalez, and R. B. Gaynor. 1989. Role of Sp1-binding domains in *in vivo* transcriptional regulation of the human immunodeficiency virus type 1 long terminal repeat. *J. Virol.* **63**:2585–2591.

[54] Hart, C. E., M. J. Saltrelli, J. C. Galphin, and G. Schochetman. 1995. A human chromosome 12-associated 83-kilodalton cellular protein specifically binds to the loop region of human immunodeficiency virus type 1 trans-activation response element RNA. *J. Virol.* **69**:6593–6599.

[55] Hattori, K., P. Angel, M. LeBeau, and M. Karin. 1988. Structure and chromosomal localization of the functional intronless human JUN protooncogene. *Proc. Natl. Acad. Sci. USA* **85**:9148–9152.

[56] Hermann, C. H., and A. P. Rice. 1993. Specific interaction of the human immunodeficiency virus Tat protein with a cellular protein kinase. *Virology* **197**:601–608.

[57] Huang, L. M., and K.-T. Jeang. 1993. Increased spacing between Sp1 and TATAA renders HIV-1 replication defective: Implication for Tat function. *J. Virol.* **67**:6937–6944.

[58] Huang, L. M., A. Joshi, R. Willey, J. Orenstein, and K. T. Jeang. 1994. Human immunodeficiency viruses regulated by alternative trans-activators: genetic evidence for a novel non-transcriptional function of Tat in virion infectivity. *EMBO J.* **13**:2886–2896.

[59] Jeang, K. T., R. Chun, N. H. Lin, A. Gatignol, C. G. Glabe, and H. Fan. 1993. In vitro and *in vivo* binding of human immunodeficiency virus type 1 Tat protein and Sp1 transcription factor. *J. Virol.* **67**:6224–6233.

[60] Jones, K. A., J. T. Kadonaga, P. A. Luciw, and R. Tjian. 1986. Activation of the AIDS retrovirus promoter by the cellular transcriptin factor, Sp1. *Science* **232**:755–759.

[61] Jones, K., P. Luciw, and N. Duchange. 1988. Structural arrangements of transcription control domains within the 5′ untranslated leader regions of HIV-1 and HIV-2 promoters. *Genes Dev.* **2**:1101–1114.

[62] Jones, K. A., and B. M. Peterlin. 1994. Control of RNA initiation and elongation at the HIV-1 promoter. *Annu. Rev. Biochem.* **63**:717–743.

[63] Jones, K. A., and R. Tjian. 1985. Sp1 binds to promoter sequences and activates herpes simplex virus "immediate-early" gene transcription *in vitro*. *Nature* **317**:179–182.

[64] Kaczmarski, W., and S. A. Khan. 1993. Lupus autoantigen Ku protein binds HIV-1 TAR RNA *in vitro*. *Biochem. Biophys. Res. Communications* **196**:935–942.

[65] Kadonaga, J. T., K. R. Carner, F. R. Masiarz, and R. Tjian. 1987. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* **51**:1079–1090.

[66] Kamine, J., B. Elangovan, T. Subramanian, D. Coleman, and G. Chinnadurai. 1996. Identification of a cellular protein that specifically interacts with the essential cysteine region of the HIV-1 Tat transactivator. *Virology* **216**:357–366.

[67] Kamine, J., T. Subramanian, and G. Chinnadurai. 1991. Sp1-dependent activation of a synthetic promoter by human immunodeficiency virus type I Tat protein. *Proc. Natl. Acad. Sci. USA* **88**:8510–8514.

[68] Kashanchi, F., G. Piras, M. F. Radonovich, J. F. Duvall, A. Fattaey, C.-M. Chiang, R. G. Roeder, and J. N. Brady. 1994. Direct interaction of human TFIID with the HIV-1 transactivator Tat. *Nature* **367**:295–299.

[69] Kato, H., M. Horikoshi, and R. G. Roeder. 1991. Repression of HIV-1 transcription by a cellular protein. *Science* **251**:1476–1479.

[70] Kenan, D. J., C. C. Query, and J. D. Keene. 1991. RNA recognition: towards identifying determinants of specificity. *Trends Biochem. Sci.* **16**:214–220.

[71] Kieran, M., V. Blank, F. Logeat, J. Vandekerckhove, F. Lottspeich, O. LeBail, M. B. Urban, P. Kourilsky, P. A. Baeuerle, and A. Israel. 1990. The DNA binding subunit of NF-kB is identical to factor KBF-1 and homologous to the rel oncogene product. *Cell* **44**:261–272.

[72] Kim, J., F. Gonzalez-Scarano, S. Zeichner, and J. Alwine. 1993. Replication of type 1 human immunodeficiency viruses containing linker substitution mutations in the –201 to –130 region of the long terminal repeat. *J. Virol.* **67**:1658–1662.

[73] Krainer, A. R., A. Mayeda, D. Kozak, and G. Binns. 1991. Functional expression of cloned human splicing factor SF2: Homology to RNA-binding proteins, U1, 70k, and Drosophila splicing factor. *Cell* **66**:383–394.

[74] Kubota, S., Y. Adachi, T. D. Copeland, and S. Oroszlan. 1995. Binding of human prothymosin alpha to the leucine-motif/activation domains of HTLV-I Rex and HIV-1 Rev. *Eur. J. Biochem* **233**:48–54.

[75] Lee, C. H., B. Leung, M. A. Lemmon, J. Zheng, D. Cowburn, J. Kuriyan, and K. Saksela. 1995. A single amino acid in the SH3 domain of Hck determines its high affinity and specificity in binding to HIV-1 Nef protein. *EMBO J.* **14**:5006–5015.

[76] Lee, J.-S., K. M. Galvin, and Y. Shi. 1993. Evidence for physical interaction between the zinc-finger transcription factor YY1 and Sp1. *Proc. Natl. Acad. Sci. USA* **90**:6145–6149.

[77] Lewin, B. 1990. Commitment and activation at PolII promoters: a tail of protein-protein interactions. *Cell* **61**:1161–1164.

[78] Li, C., C. Lai, D. S. Sigman, and R. Gaynor. 1991. Cloning of a cellular factor, interleukin binding factor, that binds to NFAT-like motifs in the human immunodeficiency virus long terminal repeat. *Proc. Natl. Acad. Sci. USA* **88**:7739–7743.

[79] Liang, P., and A. B. Pardee. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**:967–971.

[80] Liou, H. C., G. P. Nolan, S. Ghosh, F. Fujita, and D. Baltimore. 1992. The NF-kB p50 precursor, p105 contains an internal IkB-like inhibitor that preferentially inhibits p50. *EMBO J.* **11**:3003–3009.

[81] Lu, Y., N. Touzjian, M. Stenzel, T. Dorfman, J. G. Sodroski, and W. A. Haseltine. 1990. Identification of cis-acting repressive sequences within the negative regulatory element of human immunodeficiency virus type 1. *J. Virol.* **64**:5226–5229.

[82] Luban, J., K. L. Bossolt, E. K. Franke, K. V. Ganjam, and S. P. Goff. 1993. Human immunodeficiency virus type 1 gag protein binds to cyclophilins A and B. *Cell* **73**:1067–1078.

[83] Luo, T., and J. V. Garcia. 1996. The association of Nef with a cellular serine/threonine kinase and its enhancement of infectivity are viral isolate dependent. *J. Virol.* **70**:6493–6496.

[84] Luo, Y., H. Yu, and B. M. Peterlin. 1994. Cellular protein modulates effects of human immunodeficiency virus type 1 Rev. *J. Virol.* **68**:3850–3856.

[85] Maekewa, T., T. Sudo, M. Kurimote, and S. Ishii. 1991. USF-related transcription factor, HIV-TF1, stimulates transcription of human immunodeficiency virus-1. *Nucl. Acids. Res.* **19**:4689–4694.

[86] Mallardo, M., E. Dragonetti, F. Baldassarre, C. Ambrosino, G. Scala, and I. Quinto. 1996. An NF-kB site in the 5′-untranslated leader regiion of the human immunodeficiency virus type 1 enhances the viral expression in response to NF-kB-activating stimuli. *J. Biol. Chem.* **271**:20820–20827.

[87] Marciniak, R. A., M. A. Garcia-Blanco, and P. A. Sharp. 1990. Identification and characterization of a HeLa nuclear protein that specifically binds to the trans-activation-response (TAR) element of human immunodeficiency virus. *Proc. Natl. Acad. Sci. USA* **87**:3642–3646.

[88] Margolis, D. M., M. Somasundaran, and M. R. Green. 1994. Human transcription factor YY1 represses human immunodeficiency virus type 1 transcription and virion production. *J. Virol.* **68**:905–910.

[89] Mattaj, I. W. 1993. RNA recognition: A family matter? *Cell* **73**:837–840.

[90] Mavankal, G., O. S. H. Ignatius, H. Oliver, D. Sigman, and R. B. Gaynor. 1996. Human immunodeficiency virus type 1 and 2 Tat proteins specifically interact with RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **93**:2089–2094.

[91] McCaffrey, P., C. Luo, T. Kerppola, J. Jain, T. Badalian, A. Ho Burgeon, E., W. Lane, J. Lambert, T. Curran, G. Verdine, A. Rao, and P. G. Hogan. 1993. Isoation of the cyclosporin-sensitive T cell transcription factor NFATp. *Science* **262**:750–754.

[92] McCormack, S., D. Thomis, and C. Samuel. 1992. Mechanism of interferon action: identification of a RNA-binding domain within the N-terminal region of the human RNA dependent P1/eIF-2 alpha protein kinase. *Virology* **188**:47–56.

[93] McMillan, N. A., R. F. Chun, D. P. Siderovski, J. Galabru, W. M. Toone, C. E. Samuel, T. W. Mak, A. G. Hovanessian, K. T. Jeang, and B. R. Williams. 1995. HIV-1 Tat directly interacts with the interferon-induced, double-stranded RNA-dependent kinase, PKR. *Virology* **213**:413–424.

[94] Meyer, B. E., J. L. Meinkoth, and M. H. Malim. 1996. Nuclear transport of human immunodeficiency virus type 1, visna virus, and equine infectious anemia virus Rev proteins: identification of a family of transferable nuclear export signals. *J. Virol.* **70**:2350–2359.

[95] Miyazaki, Y., T. Takamatsu, T. Nosaka, S. Fujita, T. E. Martin, and M. Hatanaka. 1995. The cytotoxicity of human immunodeficiency virus type 1 Rev: implications for its interaction with the nucleolar protein B23. *Exp. Cell. Res.* **219**:93–101.

[96] Muesing, M., D. Smith, and D. Capon. 1987. Regulation of mRNA accumulation by human immunodeficiency virus trans-activator protein. *Cell* **48**:691–701.

[97] Nabel, G., and D. Baltimore. 1987. An inducible transcription factor activates expression of human immunodeficiency virus in T cells. *Nature* **326**:711–713.

[98] Nelbrock, P., P. Dillon, A. Perkins, and C. A. Rosen. 1990. A cDNA for a protein that interacts with the human immunodeficiency virus Tat transactivator. *Science* **248**:1650–1653.

[99] Nikolakaki, E., G. Simos, S. D. Georgatos, and T. Giannakouros. 1996. A nuclear envelope-associated kinase phosphorylates arginine-serine motifs and modulates interactions between the lamin B receptor and other nuclear proteins. *J. Biol. Chem.* **271**:8365–8372.

[100] Nolan, G., S. Ghosh, H. C. Liou, P. Tempst, and D. Baltimore. 1991. DNA-binding and IkB inhibition of the cloned p65 subunit of NF-kB, a rel-related polypeptide. *Cell* **64**:961–969.

[101] Nunn, M. F., and J. W. Marsh. 1996. Human immunodeficiency virus type 1 Nef associates with a member of the p21-activated kinase family. *J. Virol.* **70**:6157–6161.

[102] Ohana, B., P. A. Moore, S. M. Ruben, C. D. Southgate, M. R. Green, and C. A. Rosen. 1993. The type1 human immunodeficiency virus Tat binding protein is a transcriptional activator belonging to an additional family of evolutionarily conserved genes. *Proc. Natl. Acad. Sci. USA* **90**:138–142.

[103] Park, H., M. V. Davis, J. O. Langland, H. Chang, Y. Nam, J. Tartaglia, E. Paoletti, B. L. Jacobs, R. Kaufman, and S. Venkatesan. 1994. TAR RNA-binding protein is an inhibitor of the interferon-induced protein kinase PKR. *Proc. Natl. Acad. Sci. USA* **91**:4713–4717.

[104] Pazin, M. J., P. L. Sheridan, K. Cannon, Z. Cao, J. G. Keck, J. T. Kadonaga, and K. A. Jones. 1996. NF-kappa B-mediated chromatin reconfiguration and transcriptional activation of the HIV-1 enhancer *in vitro*. *Genes Dev.* **10**:37–49.

[105] Pengue, G., A. Caputo, C. Rossi, G. Barbanti-Brodano, and L. Lania. 1995. Transcriptional silencing of human immunodeficiency virus type 1 long terminal repeat-driven gene expression by the Kruppel-associated box repressor domain targeted to the transactivating response element. *J. Virol.* **69**:6577–6580.

[106] Pfeifer, K., B. Weiler, D. Ugarkovic, M. Bachmann, H. C. Schroder, and W. E. G. Muller. 1991. Evidence for a direct interaction of Rev protein with nuclear envelope mRNA-translocation system. *Eur. J. Biochem* **199**:53–64.

[107] Rasty, S., P. Thatikunta, J. Gordon, K. Khalili, S. Amini, and J. C. Glorioso. 1996. Human immunodeficiency virus tat gene transfer to the murine central nervous system using a replication-defective herpes simplex virus vector stimulates transforming growth factor beta 1 gene expression. *Proc. Natl. Acad. Sci. USA* **93**:6073–6078.

[108] Reddy, T. R., M. Suhasini, J. Rappaport, D. J. Looney, G. Kraus, and F. Wong-Staal. 1995. Molecular cloning and characterization of a TAR-binding nuclear factor from T cells. *AIDS. Res. Hum. Retroviruses* **11**:663–669.

[109] Ross, E., A. Buckler-White, A. B. Rabson, G. Englund, and M. A. Martin. 1991. Contribution of NF-kB and Sp1 binding motifs to the replicative capacity of human immunodeficiency virus type 1: distinct patterns of viral growth are determined by T-cell types. *J. Virol.* **65**:4350–4358.

[110] Rossi, F., A. Gallina, and G. Milanesi. 1996. Nef-CD4 physical interaction sensed with the yeast two-hybrid system. *Virology* **217**:397–403.

[111] Rounseville, M. P., and A. Kumar. 1992. Binding of a host cell nuclear protein to the stem region of human immunodeficiency virus type 1 trans-activation-responsive RNA. *J. Virol.* **66**:1688–1694.

[112] Roy, S., M. Agy, A. Horvenessian, N. Sonenberg, and M. Katze. 1991. The integrity of the stem structure of human immunodeficiency virus type 1 Tat-responsive sequence RNA is required for interaction witht he interferon-induced 68,000–Mr protein kinase. *J. Virol.* **65**:632–640.

[113] Ruhl, M., M. Himmelspach, G. M. Bahr, F. Hammerschmid, H. Jaksche, B. Wolff, H. Aschauer, G. K. Farrington, H. Probst, D. Bevec, *et al.* 1993. Eukaryotic initiation factor 5A is a cellular target of the human immunodeficiency virus type 1 Rev activation domain mediating trans-activation. *J. Cell Biol.* 123(6 Pt 1):1309–1320.

[114] Ruhl, M., M. Himmelspach, G. M. Bahu, F. Mammerschmid, B. Jaksche, H. Wolffe, H. Aschauer, G. K. Farrington, H. Probst, D. Bevec, and J. Hauber. 1993. Eukaryotic initiation factor 5A is a acellular target of the human immunodeficiency virus type 1 rev activation domain mediating trans-activation. *J. Cell Biol.* **123**:1309–1320.

[115] Saksela, K., G. Cheng, and D. Baltimore. 1995. Proline-rich (PxxP) motifs in HIV-1 Nef bind to SH3 domains of a subset of Src kinases and are required for the enhanced growth of Nef+ viruses but not for down-regulation of CD4. *EMBO J.* **14**:484–491.

[116] Salghetti, S., R. Mariani, and J. Skowronski. 1995. Human immunodeficiency virus type 1 Nef and p56lck protein-tyrosine kinase interact with a common element in CD4 cytoplasmic tail. *Proc. Natl. Acad. Sci. USA* **92**:349–353.

[117] Sawai, E. T., A. Baur, H. Struble, B. M. Peterlin, J. A. Levy, and C. Cheng-Mayer. 1994. Human immunodeficiency virus type 1 Nef associates with a cellular serine kinase in T lymphocytes. *Proc. Natl. Acad. Sci. USA* **91**:1539–1543.

[118] Scala, G., M. R. Ruocco, C. Ambrosino, M. Mallardo, V. Giordano Baldassarre, F., E. Dragonetti, I. Quinto, and S. Venuta. 1994. The expression of the interleukin 6 gene is induced by the human immunodeficiency virus 1 tat protein. *J. Exp. Med.* **179**:961–971.

[119] Seeler, J. S., C. Muchardt, A. Suessle, and R. Gaynor. 1994. Transcription factor PRDII-BF1 activates human immunodeficiency virus type 1 gene expression. *J. Virol.* **68**:1002–1009.

[120] Seto, E., Y. Shi, and T. Shenk. 1991. YY1 is an initiator sequence-binding protein that directs and activates transcription *in vitro*. *Nature* **354**:241–245.

[121] Sharma, V., M. Xu, L. M. Ritter, and N. M. Wilkie. 1996. HIV-1 tat induces the expression of a new hematopoietic cell-specific transcription factor and downregulates MIP-1 alpha gene expression in activated T-cells. *Biochem. Biophys. Res. Commun.* **223**:526–533.

[122] Shaw, D., and H. L. Ennis. 1993. Molecular cloning and developmental regulation of dictyostelium discoiceum homologues of the human and yeast HIV1 tat-binding protein. *Biochem. Biophys. Res. Commun.* **193**:1291–1296.

[123] Shaw, J., P. Utz, D. Durand, J. Tooole, E. Emmel, and G. R. Crabtree. 1988. Identification of a putative regulator of early T cell activation genes. *Science* **244**:202–205.

[124] Sheldon, M., R. Ratnasabapathy, and N. Hernandez. 1993. Characterization of the inducer of short transcripts, a human immunodeficiency virus type 1 transcriptional element that activates the synthesis of short transcripts. *Mol. Cell. Biol.* **13**:1251–1263.

[125] Sheline, C. T., L. H. Milocco, and K. A. Jones. 1991. Two distinct nuclear transcription factors recognize loop and bulge residues of the HIV-1 TAR RNA hairpin. *Genes Dev.* **5**:2508–2520.

[126] Sheridan, P. L., M. Schorpp, M. L. Voz, and K. A. Jones. 1995. Cloning of an SNF2/SWI2-related protein that binds specifically to the SPH motifs of the SV40 enhancer and to the HIV-1 promoter. *J. Biol. Chem.* **270**:4575–4587.

[127] Shibuya, H., K. Irie, J. Ninomiya-Tsuji, M. Goebl, T. Taniguchi, and K. Matsumot. 1992. New human gene encoding a positive modulator of HIV Tat-mediated transactivation. *Nature* **357**:700–702.

[128] Shukla, R. R., P. L. Kimmel, and A. Kumar. 1994. Human immunodeficiency virus type 1 Rev-responsive element RNA binds to host cell-specific proteins. *J. Virol.* **68**:2224–2229.

[129] Sorbara, L. R., F. Maldarelli, G. Chamoun, B. Schiulling, S. Chokekijcahi, L. Staudt, H. Mitsuya, I. A. Simpson, and S. L. Zeichner. 1996. Human immunodficiency virus type 1 infection of H9 cells induces increased glucose transporter expression. *J. Virol.* **70**:7275–7279.

[130] Stutz, F., M. Neville, and M. Rosbash. 1995. Identification of a novel nuclear pore-associated protein as a functional target of the HIV-1 Rev protein in yeast. *Cell* **82**:495–506.

[131] Svitkin, Y. V., A. Pause, and N. Sonenberg. 1994. La autoantigen alleviates translational repression by the 5′ leader sequence of the human immunodeficiency virus type 1 mRNA. *J. Virol.* **68**;7001–7007.

[132] Swaffield, J., J. Bromberg, and S. A. Johnston. 1992. Alterations in a yeast protein resembling HIV Tat-binding protein relieve requirement for an acidic activation in GAL4. *Nature* **357**:698–700.

[133] Szebeni, A., J. E. Herrera, and M. O. Olson. 1995. Interaction of nucleolar protein B23 with peptides related to nuclear localization signals. *Biochemistry* **34**:8037–8042.

[134] Tange, T. O., T. H. Jensen, and J. Kjems. 1996. *In vitro* interaction between human immunodeficiency virus type 1 Rev protein and splicing factor ASF/SF2-associated protein, p32. *J. Biol. Chem.* **271**:10066–10072.

[135] Usheva, A., and T. Shenk. 1994. TATA-binding protein-independent initiation: YY1, TFIIB, and RNA polymerase II direct basal transcription on supercoiled template DNA. *Cell* **76**:1115–1121.

[136] Vaishnav, Y. N., M. Vaishnav, and F. Wong-Staal. 1991. Identification of a nuclear factor that specifically binds to the Rev-responsive element (RRE) of human immunodeficiency virus type 1. *New Biologist* **3**:142–150.

[137] vanStraaten, F., R. Mueller, T. Curran, C. P. vanBeveren, and I. M. Verma. 1983. Complete nucleotide sequence of a human c-onc gene: deduced amno acid sequence of the human c-fos protein. *Proc. Natl. Acad. Sci. USA* **80**:3183–3187.

[138] Vogel, B., S.-J. Lee, A. Hildebrand, W. Craig, M. Pierschbacher, F. Wong-Staal, and E. Ruoslahti. 1993. A novel integrin specificity exemplified by binding of the avb5 integrin to the basic domain of the HIV tat protein and vitronectin. *J. Cell Biol.* **121**:461–468.

[139] Wang, Z., G. F. Morris, A. P. Rice, W. Xiong, and C. B. Morris. 1996. Wild-type and transactivation-defective mutants of human immunodeficiency virus type 1 Tat protein bind human TATA-binding protein *in vitro*. *J. Acquir. Immune. Defic. Syndr. Hum. Retrovirol.* **12**:128–138.

[140] Waterman, M. L., and K. A. Jones. 1990. Purification of TCF-1a, a T-cell-specific transcription factor that activates the T-cell receptor Ca gene enhancer in a context-dependent manner. *New Biol.* **2**:621–636.

[141] Week, B., K. Desai, P. Loewenstein, M. Klotman, P. Klotman, M. Green, and H. K. Kleinman. 1993. Identification of a novel cell attachment domain in the HIV-1 Tat protein and its 90-kDa cell surface binding protein. *J. Biol. Chem.* **268**:5279–5284.

[142] Wen, W., J. L. Meinkoth, R. Y. Tsien, and S. S. Taylor. 1995. Identification of a signal for rapid export of proteins from the nucleus. *Cell* **82**:463–473.

[143] Westendorp, M. O., M. Li-Weber, R. W. Frank, and P. H. Krammer. 1994. Human immunodeficiency virus type 1 Tat upregulates interleukin-2 secretion in activated T cells. *J. Virol.* **68**:4177–4185.

[144] Wu, F., J. Garcia, D. Sigman, and R. Gaynor. 1991. tat regulates binding of the human immunodeficiency virus trans-activating region RNA loop-binding protein TRP-185. *Genes Dev.* **5**:2128–2140.

[145] Wu-Baer, F., W. S. Lane, and R. B. Gaynor. 1996. Identification of a group of cellular cofactors that stimulate the binding of RNA polymerase II and TRP-185 to human immunodeficiency virus 1 TAR RNA. *J. Biol. Chem.* **271**:4201–4208.

[146] Wu-Baer, F., D. Sigman, and R. B. Gaynor. 1995. Specific binding of RNA polymerase II to the human immunodeficiency virus trans-activating region RNA is regulated by cellular cofactors and Tat. *Proc. Natl. Acad. Sci. USA* **92**:7153–7157.

[147] Xu, Y., T. R. Reddy, W. H. Fischer, and F. Wong-Staal. 1996. A novel hnRNP specifically interacts with HIV-1 RRE RNA. *J. Biomed. Sci* **3**:82–91.

[148] Yaseen, N. R., B. Maizel, F. Wang, and S. Sharma. 1993. Comparative analysis of NFAT (nuclear factor of activated T cells) complex in human T and B lymphocytes. *J. Biol. Chem.* **268**:14285–14293.

[149] Yoon, J.-B., G. Li, and R. Roeder. 1994. Characterization of a familyh of related cellular transcription factors which can modualte human immunodeficiency virus type 1 transcription *in vitro*. *Mol. Cell. Biol.* **14**:1776–1785.

[150] Zauli, G., D. Gibellini, A. Caputo, A. Bassini, M. Negrini, M. Monne, M. Mazzoni, and S. Capitani. 1995. The human immunodeficiency virus type-1 Tat protein upregulates Bcl-2 gene expression in Jurkat T-cell lines and primary peripheral blood mononuclear cells. *Blood* **86**:3823–3834.

[151] Zauli, G., M. Previati, E. Caramelli, A. Bassini, E. Falcieri, D. Gibellini, L. Bertolaso, D. Bosco, I. Robuffo, and S. Capitani. 1995. Exogenous human immunodeficiency virus type-1 Tat protein selectively stimulates a phosphatidylinositol-specific phospholipase C nuclear pathway in the Jurkat T cell line. *Eur. J. Immunol.* **25**:2695–2700.

[152] Zhou, Q., and P. A. Sharp. 1996. Tat-SF1: Cofactor for stimulation of transcriptional elongation by HIV-1 Tat. *Science* **274**:605–609.

[153] Zimmermann, K., M. Dobrovnik, C. Ballaun, D. Bevec, J. Hauber, and E. Bohnlein. 1991. trans-Activation of the HIV-1 LTR by the HIV-1 Tat and HTLV-I Tax proteins is mediated by different cis-acting sequences. *Virology* **182**:874–878.

HIV and Molecular Mimicry

Gerald Myers and Hong Lu

MS K710, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

For more than three decades, molecular mimicry has been considered a contributing factor in viral pathogenesis. The phenomenon comes about when viral antigens share epitopes with host cell proteins. Because tolerance is thought never to be complete, the mimicking microbe can induce anti-self antibodies, a state of autoimmunity, with resulting immunopathogenesis. This mode of attack seems preeminently possible with immunosuppressive viruses such as HSV-1 and HIV; it is also thought likely when cross-species-induced pathogenesis is involved. For some general discussions of molecular mimicry and autoimmunity brought about by viruses, see [1,2].

In the context of AIDS pathogenesis, some have argued over the years that HIV mimicry could be the crucial factor in the disease expression, for example [3–7]. If this be so, clearly the chances for a safe and successful HIV vaccine are seriously weakened [8,9]. Thus it is of considerable interest to examine mimicry claims that have been brought forward over the last decade of AIDS research. In this section, we offer a general critique of mimicry claims in so far as they have been based upon sequence similarities; experimentally-derived data will be cited on occasion, however the focus is primarily upon amino acid sequences and peptide structures. For his reason, this section appears in both the genetic and immunologic sections of the 1996 database compendium.

A. PROLEGOMENA

An important distinction that is often blurred in the literature is the distinction between *similarity* and *homology*. The correct meaning of similarity, according to the prevailing school of thought, is as an empirical relationship that can be quantified; whereas homology is an inference, a qualitative prediction about common ancestry [10–12]. Accordingly, it is impossible to prove that any two sequences are not homologous [12]. Two sequences can be similar without being homologous as a result of evolutionary convergence, and we imagine this to be the basis for viral molecular mimicry.

From this point of view, it is careless to speak of homologies in the context of mimicry, which by definition implies similarity. A pair of sequences, or a pair of structures, can be 10%, 30%, 95%... similar, but they are either homologous or they are not; philosophers speak of the first kind of judgement as determinate, whereas the judgement about homology is indeterminate. Finally, the term similarity usually makes allowances for equivalencies as well as identities: the amino acids serine and threonine may be regarded as equivalent in certain positions in a character string, and a scoring matrix would assign partial credit to one substituted for the other.

Sequences with similarity of 50% or more will typically have similar structures, although it is now well-known that two identical pentapeptides could have different shapes or presentations in space [13]. On the other hand, sequences that are marginally similar, say 30%, may have equivalent structures, thus it is not a trivial matter to show that two dissimilar sequences are also structurally dissimilar. While the emphasis herein will be upon sequence similarities, the complex relationship between sequence space and shape space must be held in mind throughout.

Moreover, it may be the case that for effective breakthrough of tolerance a mimicking epitope of a virus will not be perfectly identical to the host "self" protein. For example, an experimentally-verified mimicry between rabbit myelin basic protein (MBP) and hepatitis B virus polymerase (HBVP) is [2]:

```
MBP    TTHYGSLPQK
         YGSLPQ
HBVP   IGCYGSLPQE
```

In other cases, however, it is known that a single amino acid replacement can alter binding of an antibody to a protein [14].

We understand that some epitopes are discontinuous. Mimicry involving these epitopes will not be addressed herein. Antigen presentation is yet another complexity that will be sidestepped.

It is tempting to evaluate mimicry claims by invoking secondary structure prediction algorithms such as SOPM, discussed in an accompanying Part III section of the 1996 compendium. But we must also keep in mind long-range interactions, oligomerization, and protein residue modification (phosphorylation, glycosylation, etc.), all of which can invalidate judgements based on the primary sequence comparison and secondary structure prediction. Furthermore, we must ask whether the putative mimicking epitope on the viral antigen represents a host "self" peptide that is similarly presented on a surface [15].

It would appear from these manifold complications that mimicry can only be satisfactorily critiqued from an experimental standpoint, and indeed Oldstone emphasizes testability in his review of viral mimicry (see figure 1 in reference 2, for example). Be that as it may, untested mimicry hypotheses fill the literature of HIV—some have gained strength from time alone—and therefore critique at the sequence level might provide some insight in the absence of experimental test.

It occurs to us, in particular, that many mimicry claims made on the basis of suspected sequence similarities in the 1980's should be revisited today when the database is orders of magnitude larger. If mimicry is a pervasive factor in AIDS pathogenesis, then the sequence/structural similarity would probably be observed across many diverse variants and homologous types. In 1987, for example, a borderline similarity between HIV envelope and a molecule called neuroleukin was claimed; the story was an intriguing one in so far as it offered a basis for AIDS neuropathogenesis [16]. An experiment performed with an SIV envelope accompanied the sequence claim, but unfortunately the authors neglected to also assess the similarity between the SIV envelope and the so-called neuroleukin molecule, which turned out to be weaker than borderline. It was subsequently shown that the putative neuroleukin was, in fact, a commonplace enzyme; yet the neuroleukin hypothesis for AIDS-related neuropathy persisted in the literature after the sequences and experiment had been fully critiqued [3].

In the following section, we shall first discuss the potential contribution of compositional similarities to mimicry. Subsequent sections will include, among other things, a catalog of mimicry claims related to HIV envelope protein. At a later time, this discussion might be expanded to cover other HIV proteins, especially Gag and Nef epitopes, which are also thought to be involved in mimicry (see Part III section on Alignments, Database Searching, and Structure Prediction).

B. COMPOSITIONAL CONSIDERATIONS

One strategy for discovering similarities between HIV antigens and host proteins is to start with the amino acid composition of the former. Lentiviruses have skewed base compositions (high A, low C) and unusual codon preferences that result in significant compositional differences from what is generally seen in the PIR and Swiss-Prot protein databases [17]. HIV has a relatively high tryptophan content, for example, and tryptophan is one of the least frequently found amino acids in eukaryotes and most viruses. Polar residues are relatively high in HIV. Starting from this fact, Fitzgerald and coworkers cataloged microbial sequences that shared with HIV the unusual base composition [18]. Of these, most were bacterial virulence factors, *i.e.*, surface antigens, hence the authors speculated about similarities of immune dysfunction brought about by such diverse pathogens as protozoans, Pneumocystis, Treponema, and HIV. While the Fitzgerald study does not direct itself to HIV mimicry of host antigens (so much as it is concerned with antigen switching), the overall premise is heuristic: mimicry probably begins with gross similarities—similar charge, similar presentation, similar extent of glycosylation, etc. PROPSEARCH (http://www.heidelberg.de/aaa.html) and AACompIdent (http://expasy.chuge.ch/ch2d/aacompi.html) are two programs accessible through Internet for identifying a protein from its amino acid composition.

Taking a more focused approach from what has just been described for HIV and compositional similarities, Douvas and Sobelman conducted sweeping sequence comparisons involving three preselected sequence sets: a set of nuclear antigens, some of which are known to be involved in autoimmune disorders; a set of viral proteins, some of which are related to immunosuppression; and a set of 41 control proteins not known to have autoimmune significance [19]. Pentameric and hexameric similarities were

found to the human nuclear antigens in both the viral and control sets, but an excess of "hits" was observed for viral proteins and two nuclear antigens—the 70 kDa component of RNP particles, which is characteristically involved in mixed connective tissue disease (MCTD), and CENP-B, a centromere protein which is associated with scleroderma. Both proteins are highly hydrophilic. Several viruses, HIV-1, HSV-1, EBV, SRV-1, and CMV, were implicated. The proposed matches to HIV were to both gp120 and gp41, and because the strongest instances involved the 70 kDa splicing protein, the authors followed that up in [20].

The similarity between HIV-1 gp41 and the hydrophilic COOH-terminus of the 70 kDa ribonucleoprotein component has undeniable compositional implications, as is evident below

```
HMM Consensus (gp41)    GPDRPEGIEEEGGEQDRDR
                        GPD P G EE G ++DR R
U1-snRNP 70 kDa         GPDGPDGPEEKGRDRDRER
```

A comparable situation of sequence compositional bias is found with a stretch of the HIV/SIV Nef protein and the CENP-B centromere antigen studied in [19]:

```
HMM Consensus (Nef)     DDWDDEDEEEVGFP
                        DD D+ED +EV +P
CENP-B                  DDDDEEDGDEVPVP
```

These alignments were taken from a Smith-Waterman (S-W) search using an HMM-generated consensus for HIVs, as described in an accompanying section of Part III of the 1996 compendium. The log-odds score in each case (in bits) is about 1, which implies a two-fold greater than even probability that the similarity does not obtain from chance alone; as a rule a score should be 20 or more to have high specificity. The similarities and the marginal scores seen in these alignments stem from the low complexity (relative monotony) of the sequences involved.

The HMM Env consensus, which was generated from about 400 HIV and SIV sequences, extends the observations of [19] and [20] to primate immunodeficiency viral sequences in general, but it also reveals the ease with which such a weak match can be found: for comparison, the similarity score between the relevant region of gp41 and the *homologous* CAEV transmembrane protein is 5.5, not a high score but more than twenty-fold higher than the score implied by 1. This is not itself a judgment that the mimicry claim is false; rather, it merely points to a weakness in the sequence argument that led to, and helps sustain, the claim. Supporting the claim are the observations that: 1) the COOH-terminus of the 70 kDa can itself beget autoantibodies (which can interfere with splicing function); 2) the region of gp41 involved is known to be immunogenic; 3) similarities, however lacking in complexity, are found in clusters rather than at a single locus; 4) the authors find a 50 to 100% enhancement over control sera of binding of anti-RNP antibodies to gp41 in an ELISA assay; and 5) HIV sera show reactivity to snRNP 70 kDa in western blotting [20].

In the face of weak similarities such as those shown above, it is instructive to search a database larger than 41 control proteins to discover the ease with which matches to the 70 kDa and gp41 regions in question are encountered. Running a BLASTP search with the 70 kDa splicing protein sequence (BLAST also computes log-odds scores), more than 900 matches can be uncovered which have scores and Poisson probabilities superior to those associated with the relevant gp41 fragment. Among viruses, many represent more likely matches to 70 kDa peptides than does HIV gp41; and not all of these possibly mimicking viruses are immunosuppressive (*e.g.* human papillomaviruses). Hence, what we might call "dry mimicry" (computer-based similarity) of the 70 kDa splicing protein may be commonplace; it remains to be seen how extensive the "wet mimicry" will be.

If gp41 mimics the 70 kDa splicing protein, it might also have the potential to mimic other cellular proteins on the same basis of hydrophilicity. BLASTing the database for other matches to just the relevant C-terminal portion of the HMM consensus gp41 sequence produces an array of weak matches to several cellular proteins. One of the strongest of these borderline matches is indeed the 70

kDa RNP (Poisson probabilities between 0.07 and 0.11). This is in some contrast to the comparatively weaker score from the S-W search based on the entire HMM consensus for Env (see above).

Mimicry analyses can also be fruitfully conducted using the BLOCKS database, version 9.1, which consists of 3300 multiply aligned, gapless arrays of similarities and which features position-dependent scoring [21,22] (http://www.blocks.fhcrc.org). The snRNP family of proteins is represented in the BLOCKS database as three blocks, BL00030A,B,C, the third of which is most representative of the COOH-terminus of the 70 kDa splicing protein. One of the distinct advantages of the BLOCKS search strategy is that clusters of similarity can be identified. A search of the database using the relevant COOH-terminus of the gp41 consensus sequence did not reveal a significant match to any of the 3300 blocks. We will have reason to return to the BLOCKS search strategy below. At this point in our assessment, we have some experimental evidence in support of the claim, but the sequence arguments that underpin the claim appear to be insignificant.

Before leaving the topic of compositional bias, a brief comment about probabilities is in order. *A priori* probabilities based on the fraction 1/20 raised to some power are generally neither accurate nor useful for sequences of epitope length; they typically underestimate the likelihood of meaningless similarities, especially when compositional bias is present. In general, statistical relationships between a query sequence and target database sequences will be strongly affected by sequence length and composition, and the presumption of a normal distribution in sequence searching can distort the results [23]. Log-odds scores, used in PAM matrices, BLAST, and HMM-related database searching, are more trustworthy. For example, the HMM/S-W search strategy [24],

$$\text{score} = \log_2 \frac{P(S_i|M)}{P(S_i|R)}$$

where the alignment of each sequence in the database, S_i, is compared to both the HMM generated model, M, and a random model, R. The latter should have the same amino acid composition as the database at large and it should be as likely, *a priori*, as M. The log-odds score corrects for sequence length. With BLAST, as we have seen in a previous paragraph, Poisson probabilities are also calculated and reported alongside the log-odds score [25]. Output from a BLOCKS search includes a nomogram to assist the user in interpreting the probability of a chance similarity.

C. SEQUENCE DEGENERACY and MIMICRY

A second mimicry claim regarding HIV envelope and the 70kDa splicing protein involves the well-studied gp120 V3 loop of the former and the nucleic acid binding sequence of the latter, both of which are immunogenic [20]. In this instance, low complexity, or compositional uniformity, is not obviously involved; on the other hand, widespread variability in the V3 loop calls into question the breadth of the claim:

```
                    HIV-1 IIIB   RGPGRAFVTIGK
                  U1-snRNP 70 kDa  KPRGYAFIEYEH
HIV-1 B subtype consensus      IGPGRAFYTTGE
                               mawrkvwfar-d
                               lqlkqtlhr-eq
                               v qwssvv-ark
                               t f ggiigidg
                               f a -ryrsnkr
                               k g  qmwvkta
                               r v  xsdysnn
                               s s   tt yqh
                               y t   x  pss
                               m       xp
```

The authors in [20] took note of the similarity between the HIV-1 IIIB V3 loop and the 70 kDa splicing protein, shown in the top two lines of the alignment. Shown below them is the comparison

with the B subtype consensus sequence and in lowercase letters, in descending order of frequency, the observed replacements at the various positions in B subtype loops as catalogued in Part III of the HIV database compendium. Other HIV-1 subtypes, A, D, E, O, etc., manifest variable V3 loops and HIV-2s have a very different sequence and structure altogether. This raises the question whether mimicry occurs with select variants when it doesn't occur throughout a viral type.

We do not find that the HMM consensus for the HIV/SIV V3 loop region takes part in any significant matches in a S-W/HMM search. Furthermore, a meaningful similarity is not found in a BLOCKS database, even though the 70 kDa region of interest, the nucleic acid binding sequence, is represented as block BL00030B. On the other hand, reactivities between anti-RNP sera and the IIIB V3 antigen in an ELISA assay were found to be as great as those between HIV positive sera and the IIIB loop. Furthermore, in one MCTD patient who was infected with HIV, anti-V3 and anti-70 kDa titers were observed to vary in tandem [20]. Finally, we should note that an established case of mimicry, the rabbit myelin basic protein and the hepatitis B viral polymerase discussed above, would not yield significant BLAST similarity.

We must tentatively conclude that some serological effects are being captured, yet we must also wonder how extensive they can be given the well-documented sequence degeneracy. (It is conceivable that some of the V3 loop sequences in the database were taken from defective viruses, but it is also certainly the case that the IIIB loop sequence is not the only viable sequence.) Finally, it is simply possible that the effects seen for the IIIB V3 and the 70 kDa RNP component are not biologically significant, in spite of the cross-reactivities and some sequence similarity (however weak) [26].

HIV is one of the most degenerate (variable) microbial pathogens. The previous example suggests that hypotheses based on sequences from the earliest characterized HIV-1 strains, such as IIIB, may not be indicative of a general picture of mimicry, unless viral mimicry is thought to be sometimes particular, rather than being uniformly generic. The following table of mimicry claims in the literature involving HIV envelope and host-cell proteins sequences attempts to summarize this situation. The summary is not exhaustive, but it covers the majority of claims about mimicry involving HIV gp120 and gp41 and known host protein sequences. The table lists the original claim, reports a best individual match set (first three columns), then presents the claim in light of the variation encountered in both the viral and the cell protein (fourth column). HMM-generated consensus sequences have been taken from Part II of the 1996 database compendium; their generation is discussed in a separate section of Part III. In some instances, a consensus could be deduced for the cellular protein (fourth column).

One immediately notes that a major fraction of the HIV envelope mimicry claims involve immune-related host proteins—HLA, Ig, complement, etc. This is partly an historical artifact: many of these claims were brought forward in the 1980's when an overabundance of immunoglobulin-related sequences was present in the protein libraries, hence a database search with any query sequence was likely to stumble upon borderline matches to immune-related molecules [23,27]; and in the case of an HIV sequence query, any connection to immunopathology is suggestive. This situation was of course exacerbated by the fact that the full range of HIV variation was unappreciated at the time.

Many of the claims included in the table have been critiqued at the bench by Neurath and coworkers, who simply conclude that "immunization with gp120/gp160 is unlikely to elicit harmful autoimmune responses" [8].

gp120

Source	Protein Coordinate	Sequence	Consensus	Ref
IgG1, IgG2, IgG3	88-97	NHKPSNTKVDK	nHKPSNTKVDK	28
HIV-1/gp120(DJ258A)	55-65	DAKAYDTEVHN	dakay?tevhn	
Most-likely HIV12-SIV			DAKAYDTEVHN	
HLA DR II beta	142-151	VVSTGLIHNG	vvst?(?)li?ng	29,30
Fas antigen	275-280	VQLIRN		5
HIV-1/gp120 (IBNG)	251–260	VVSTQLLLNG	vvstqlllnG	
Most-likely HIV12-SIV			VVSTQLLLNG	
HLA DR alpha	28-40	EEHVIIQAEFYLN		29
HIV-1/gp120 (2HT596.4)	270-282	EEVVIRSANFTDN	e(?)e??irsen?tnn	
Most-likely HIV12-SIV			EEIVIRSENFTDN	
HLA class I C alleles	66-69, 79-82	KYKR, RKLR		31
HIV-1/gp120 (HXB2R)	485-487,500-503	KYK, KAKR	KYK, ?(?)akr	
Most-likely HIV12-SIV			KYK, KAKR	
Fas protein	115-124	VEINCTR		6,32
HIV-1/gp120	292-301	VEINCTR	v?rnctr	
Most-likely HIV12-SIV			VEINCTR	
CD4 receptor	60-64	SLWDQ		33,6,5
HIV-1/gb120 (IBNG)	108-112	SLWDQ	slwd?q	
Most-likely HIV12-SIV			SLWDQ	
RV glycoprotein	189-199	CDIFTNSRGKR		34,35
a-cobratoxin	30-40	CDAFCSIRGKR		34,35
a-bungarotoxin	30-40	CDAFCSSRGKV		34,35
HIV-1/gp120	164-174	FNISTSIRGKV	fn?tt??rdk?	
Most-likely HIV12-SIV			FNITTEIRDKK	
TCR alpha-chain	26-46	PILILKQMCHKVRILMCISQT		6,5,36
HIV-1/gp120	212-232	PIPIHYCAPAGFAILKCNNKT	pipihyCapagfailkcndk?	
Most-likely HIV12-SIV			PIPIHYCAPAGFAILKCNDKK	
Ig light chain V region	37-48	QQHPGKAPKLVI	QQkPGkAP??iI	37
HIV-1/gp120 (NL43)	308-328	QRGPGRAFYVTI	???gpg?(??)afy??t	
Most-likely HIV12-SIV			TIGPGQAFYATG	
70 kDa splicing protein	319-330	KPRGYAFIEYEH		20
HIV-1/gp120(LAI)	316-423	RGPGRAFVTIGK	?gpg?(??)afy?(?)tg?	
Most-likely HIV12-SIV			IGPGQAFYATED	
Ig light chain V region	29-39	LLHSDGFDYLN	LL??Dg????n	37
HIV-1/gp120 (HXB2R)	453-463	LLTRDGGNSNN	lltrdgg(??????)n???	
Most-likely HIV12-SIV			LLTRDGGDNNSTN	
Ig light chain		CSTDINGYFLF		38
HIV-1/gp120 (LAI)	450-460	CSSNITGLLLT	c?snitGlllt	
Most-likely HIV12-SIV			CSSNITGLLLT	
rheumatoid antigen (70-kDa)	488-492	GGGDM		19
HIV-1/gp120 (IBNG)	462-466	GGGDM	gggdm	
Most-likely HIV12-SIV			GGGDM	

gp41

Source	Protein Coordinate	Sequence	Consensus	Ref
HLA DR II alpha	169-183	VEHWGLDQPL	vehwgld?PL	29
HIV-1/gp41 (SF1703)	588-597	VERYLKDQQL	vErylkd?qql	
Most-likely HIV12-SIV			VERYLKDQQL	
rheumatoid antigen (70-kDa)	407-413	RDRDRDR		19,20
rheumatoid antigen (70-kDa)	524-528	RDRDR		
rheumatoid antigen (70-kDa)	542-552	RDRDRDRDRDR		
HIV-1/gp41 (SF2)	739-743	RDRDR	?drdr	
rheumatoid antigen (70-kDa)	562-566	ERGRD		
HIV-1/gp41 (ELI)	737-741	ERGRD	e?drdr	
Most-likely HIV12-SIV			ERDRDR	
rheumatoid antigen (CENP-B)	428-433	EEEGGE		19
HIV-1/gp41 (SF1703)	739-744	EEEGGE	eeeGGe	
Most-likely HIV12-SIV			EEEGGE	
HLA DR II beta	19-25	NGTERVR	ngterv?	39,40,41,3
HIV-1/ gp41 (LAI)	829-835	EGTDRVI	egtDrvi	
Most-likely HIV12-SIV			EGTDRVI	
IL-2	34-39	LEHLLL		42,43
HIV-1/gp41 (LAI)	856-861	LERILL	leraLl	
Most-likely HIV12-SIV			LERALL	

D. CONCLUDING REMARKS

Molecular mimicry involving viral peptides does occur, however the extent and specificity are not easy to establish. By the nature of the problem, short, typically hydrophilic, sequences are involved, making rigorous similarity searches difficult. At a minimum, sequence analysts must critique their claims with respect to 1) compositional factors; 2) statistical significance; and 3) sequence degeneracy. Other important considerations include secondary and tertiary structure, amino acid modifications, and insignificant cross-reactivities, that is to say in vitro cross-reactivities that have no biological meaning. Whether autoimmune complications play a major or a negligible role in AIDS pathogenesis remains to be seen.

References

[1] Oldstone,M.B.A. (1987) Molecular mimicry and autoimmune disease. *Cell* **50**:819-820.

[2] Oldstone,M.B.A. (1989) Virus induced autoimmunity: molecular mimicry as a route to autoimmune disease. *J Autoimmunity* **2** (Suppl.):187–194.

[3] Bjork,R.L. Jr. (1991) HIV-1: seven facets of functional molecular mimicry. *Immunol Lett* **28**:91–95.

[4] Habeshaw,J., Hounsell,E., and Dalgleish,A. (1992) Does the HIV envelope induce a chronic graft-versus-host-like disease? *Immunol Today* **13**:207–210.

[5] Susal,C., Kropelin,M., Daniel,V., and Opelz,G. (1993) Molecular mimicry between HIV-1 and antigen receptor molecules: a clue to the pathogenesis of AIDS. *Vox Sang* **65**:10–17.

[6] Silvestris,F., Williams, R.C., Jr., and Dammacco,F. (1995) Autoreactivity in HIV-1 infection: the role of molecular mimicry. *Clin Immunol Immunopath* **75**:197–205.

[7] Bisset, L.R. (1994) Molecular mimicry in the pathogenesis of AIDS: the HIV/MHC mycoplasma triangle. *Med Hypotheses* **43**:388–396.

[8] Neurath,A.R., Strick,N., Li,Y.-Y., and Jiang,S. (1993) Improbability of harmful autoimmune responses resulting from immunization with HIV-1 envelope glycoproteins. *AIDS Res Human Retro* **9**:1195–1208.

[9] Weber, G.F. and Cantor, H. (1993) HIV glycoprotein as a superantigen. A mechanism of autoimmunity and implications for a vaccination strategy. *Med Hypotheses* **41**:247–250.

[10] Fitch, W.M. (1973) Aspects of molecular evolution. *Annu. Rev. Genetics* **7**:343–380.

[11] Pearson,W.R. (1996) Effective protein sequence comparison. In *Computer Methods for Macromolecular Sequence Analysis* (ed. R.F. Doolittle) pp. 227–258. *Methods Enzymol.* Vol. 266. Academic Press, San Diego, 1996.

[12] Doolittle, R.F. (1987) *Of URFS and ORFS: A Primer on How to Analyze Derived Amino Acid Sequences.* University Science Books, Mill Valley, California.

[13] Rost,B. and Sander,C. (1993) Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* **232**:584–599.

[14] Ivanoff,L.A., Looney, D.J., McDanal, C., Morris, J.F., Wong-Staal, F., Langlois, A.J., Petteway, S.R. Jr., and Matthews, T.J. (1991) Alteration of HIV-1 infectivity and neutralization by a single amino acid replacement in the V3 loop domain. *AIDS Res Human Retro* **7**:595–603.

[15] Quaratino,S., Thorpe,C.J., Travers,P.J., and Londei,M. (1995) Similar antigenic surfaces, rather than sequence homology, dictate T-cell epitope molecular mimicry. *Proc Natl Acad Sci USA* **92**:10398–10402.

[16] Lee,M.R., Ho,D.D., and Gurney,M.E. (1987) Functional interaction and partial homology between human immunodeficiency virus and neuroleukin. *Science* **237**:1047–1051.

[17] Myers, G, Retroviral Sequences. (1996) In *Retroviruses*, ed. by Coffin,J. Hughes,S. and Varmus,H. Appendix I. Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY.

[18] Fitzgerald,D.J., Bronson,E.C., and Anderson,J.N. (1996) Compositional similarities between the human immunodeficiency virus and surface antigens of pathogens. *AIDS Res. Human Retroviruses* **12**:99–106.

[19] Douvas,A. and Sobelmen,S. (1991) Multiple overlapping homologies between two rheumatoid antigens and immunosuppressive viruses. *Proc Natl Acad Sci USA* **88**:6328–6332.

[20] Douvas,A. and Takehana,Y. (1994). Cross-reactivity between autoimmune anti-U1 snRNP antibodies and neutralizing epitopes of HIV-1 gp120/41. *AIDS Res Human Retro* **10**:253–262.

[21] Henikoff,S. and Henikoff,J.G. (1994) Protein family classification based on searching a database of blocks. *Genomics* **19**:97–107.

[22] Henikoff,J.G. and Henikoff,S. (1996) BLOCKS database and its applications. In *Computer Methods for Macromolecular Sequence Analysis*, ed. R.F. Doolittle, pp.88–105, Academic Press, San Diego.

[23] Altschul,S.F., Boguski,M.S., Gish,W., and Wooten,J.C. (1994) Issues in searching molecular sequence databases. *Nature Genetics* **6**:119–129.

[24] Eddy,S.R., Mitchison,G., and Durbin,R. (1995) Maximum discrimination hidden Markov models of sequence consensus. *J Comp. Biol.* **2**:9–23.

[25] Madden,T.L., Tatusov,R.L., and Zhang,J. (1996) Applications of network BLAST server. In *Computer Methods for Macromolecular Sequence Analysis*, ed. R.F. Doolittle, pp.131–141, Academic Press, San Diego.

[26] Nickerson, C., Luthra, H., and David, C. (1991) Antigenic mimicry and autoimmune diseases. *Int J Immunol* **7**:205–224.

[27] Claverie,J.-M. and States,D.J. (1993) Information enhancement methods for large scale sequence analysis. *Computers Chem.* **17**:191–201.

[28] Solder,B., Marschang,P., Wachter,H., Dierich,M.P., Nayyar,S., Lewin,I.V., and Stanworth,D.R. (1989). Anti-viral antibodies in HIV(HTLV-III) infection possess auto-antibody activity against a CH1 domain determinant in human IgG: possible immunological consequences. *Immunol. Lett.* **23**:9–19.

[29] Zaitseva,M.B., Moshnikov,S.A., Kozhich,A.T., Frolova,H.A., Makarova,O.D., Pavlikov,S.P., Sidorovich,I.G., and Brondz,B.B. (1992) Antibodies to MHC class II peptides are present in HIV-1 positive sera. *Scand. J. Immunol.* **35**:267–273.

[30] Young,J.A. (1988) HIV and HLA similarity. *Nature* **333**:215.

[31] Lopalco,L., De Santis, C., Meneveri,R., Longhi,R., Ginelli,E., Grassi,F., Siccardi,A.G., and Beretta,A. (1993) Human immunodeficiency virus type 1 gp120 C5 region mimics the HLA class I alpha1 peptide-binding domain. *Eur. J. Immunol.* **23**:2016–2021.

[32] Szawlowski,P.W., Hanke,T., and Randall,R.E. (1993) Sequence homology between HIV-1 gp120 and the apoptosis mediating protein Fas. *AIDS* **7**:1018.

[33] Zagury,J.-F., Cantalloube,H., et al. (1992) Striking identity between HIV-1 glycoprotein gp120 and its CD4 receptor [letter]. *Lancet* **340**:483–484.

[34] Bracci,L., Lozzi,L., Rustici,M., and Neri,P.(1992) Binding of HIV-1 gp120 to the nicotinic receptor. *FEBS Lett.* **311**:115–118.

[35] Neri,P., Bracci,L. Rustici,M. and Santucci,A. (1990) Sequence homology betweeen HIV gp120, rabies virus glycoprotein, and snake venom neurotoxins. Is the nicotinic acetylcholine receptor an HIV receptor? *Arch. Virol.* **114**:265–269.

[36] Kropelin,M., Susal,C., and Opelz,G. (1993) Cross-reactivity of anti-HIV with the CD4 receptor molecule. Proc. Int. Conf. AIDS, June 6-11, abstract PO-A20-0433.

[37] Metlas,R., Veljkovic,V., Paladini,R.D., and Ponger,S. (1991) Protein and DNA-sequence homologies between the V3 loop of human immunodeficiency virus type I envelope protein gp120 and immunoglobulin variable regions. *Biochem. Biophys. Res. Comm.* **179**:1056–1062.

[38] Kieber-Emmons,T., Jameson, B.A., and Morrow, W.J. (1989) The gp120-CD4 interface: structural, immunological and pathological considerations. *Biochem. Biophys. Acta* **989**:281–300.

[39] Golding,H., Shearer,G.M., Hillman,K., Lucas,P., Manischewitz,J., Zajac,R.A., Clerici,M., Gress,R.E., Boswell,R.N., and Golding,B. (1989) Common epitope in human immmunodeficiency virus (HIV) I-GP41 and HLA class II elicits immunosuppressive autoantibodies capable of contributing to immune dysfunction in HIV I-infected individuals. *J. Clin. Invest.* **83**:1430–1435.

[40] Golding,H., Robey,F.A., Gates, F.T., Linder, W., Beining, P.R., Hoffman, T., and Golding, B. (1988) Identification of homologous regions in human immunodeficiency virus I gp41 and human MHC class II beta 1 domain. Monoclonal antibodies against the gp41-derived peptide and patients' sera react with native HLA class II antigens, suggesting a role for autoimmunity in the pathogenesis of acquired immmunr deficiency syndrome. *J. Exp.Med.* **167**:914–923.

[41] Koshino,K., Tokano,Y., Hishikawa,T., Sekigawa,I., Takahashi,Y., and Hashimoto,H. (1995) Detection of antibodies to HIV-1 gp41 and HLA class II antigen-derived peptides in SLE patients. *Scand. J. Rheumatol.* **24**:288–292.

[42] Bost, K.L. and Pascual, D.W. (1988) Antibodies against a peptide sequence within the HIV envelope protein crossreacts with human interleukin-2. *Immunol. Invest.* **17**:577–586.

[43] Reiher,W.E., Blalock,J.E., and Brunck,T.K. (1986) Sequence homology between acquired immunodeficiency syndrome virus envelope protein and interleukin 2 [published erratum appears in *Proc Natl Acad Sci USA* **84**:564,1987]. *Proc. Natl. Acad. Sci. USA* **83**:9188–9192.

V

Communications

Part V of *Human Retroviruses and AIDS 1996* is divided into two sections. The first introduces ways of accessing HIV information from the World Wide Web (WWW), the second is a list of references.

HIV Information on the World Wide Web

All of the information in *Human Retroviruses and AIDS 1996*, its predecessors in 1994 and 1995, as well as numerous sequence and alignment data files, are available to the public through the World Wide Web, a system of links between Internet locations that can be navigated with graphical "browsing tools" such as Mosaic and Netscape. We are strenuously urging our subscribers to make use of our Web site. Using electronic documents and data offers many advantages over the production and use of paper documents.

- It saves paper, ink, metal, solvents, plastics, postage, fossil fuels, and time that go into the printing and distribution of hundreds of copies of this large book.
- Electronic documents that contain all the formatting and complex graphics of a printed document can now be produced and placed on the Web where users may view and print them. For example, an exact replica of this entire document, *Human Retroviruses and AIDS 1996* is available from our Web site.
- Electronic documents are readily corrected and kept up to date. As we prepared sections of this year's compendium we have been "prepublishing" them on our Web site. In this way the latest and most accurate information is always available to users.
- There is also the possibility of interaction between users and the database staff at Los Alamos. Just as documents and data can be retrieved from the Web site, so can they be posted to the site.
- Data and documents on the Web, far from being static pieces of paper, are embedded in a system that allows searching for items of interest—references, sequences, author names, etc. Our search capabilities are relatively primitive now but will become more powerful in the future.
- The Web, as its name implies, is a system of links spread throughout the world. Our Web site provides links to other sites that are useful to virologists and sequence biologists.

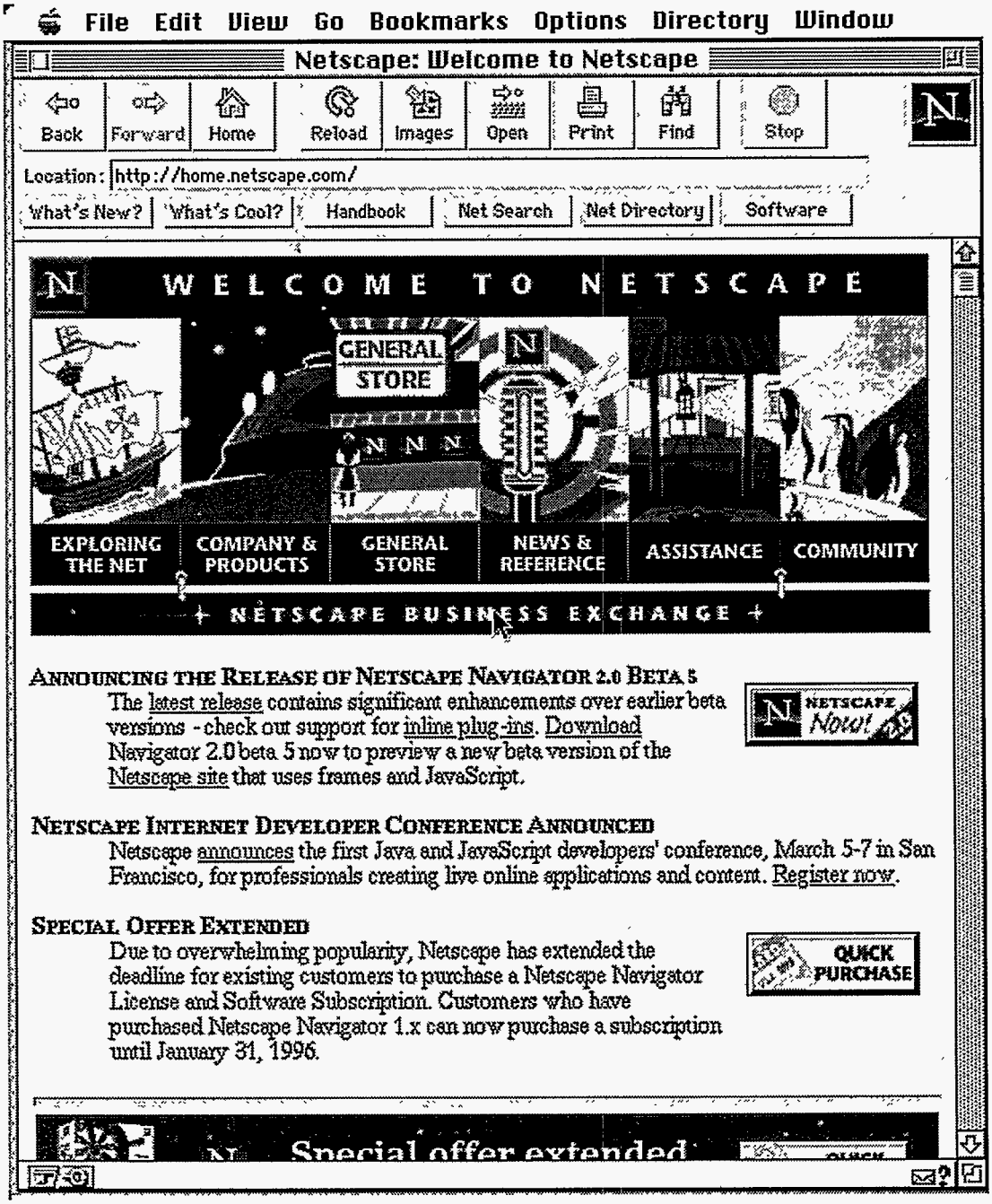
In order to access the WWW, you must have, (1) a windowing capability such as X Windows, MS Windows, or Apple Macintosh, (2) an Internet Protocol (IP) connection, and (3) a copy of a browser program such as Mosaic or Netscape for the appropriate windows environment. If you are not already using Mosaic (a free program) you can get a copy by anonymous ftp from `ftp.ncsa.uiuc.edu`. Netscape may be purchased by sending e-mail to `sales@netscape.com` or phoning 415-528-2555. Our use of Netscape in the following description of our Web site is purely illustrative and should not be construed as an endorsement of that product.

The browsing tools designed for the WWW make it easier than before to locate and acquire information on the Internet because, instead of typing arcane commands into your computer, you simply point and click at text and objects on your screen. These objects are either links to files, which cause the files to be sent to your local computer, or they are links to other machines from which files may be accessed. "Files" can mean words, pictures, sounds, or movies. The WWW address (or "URL") for the HIV database is:

`http://hiv-web.lanl.gov`

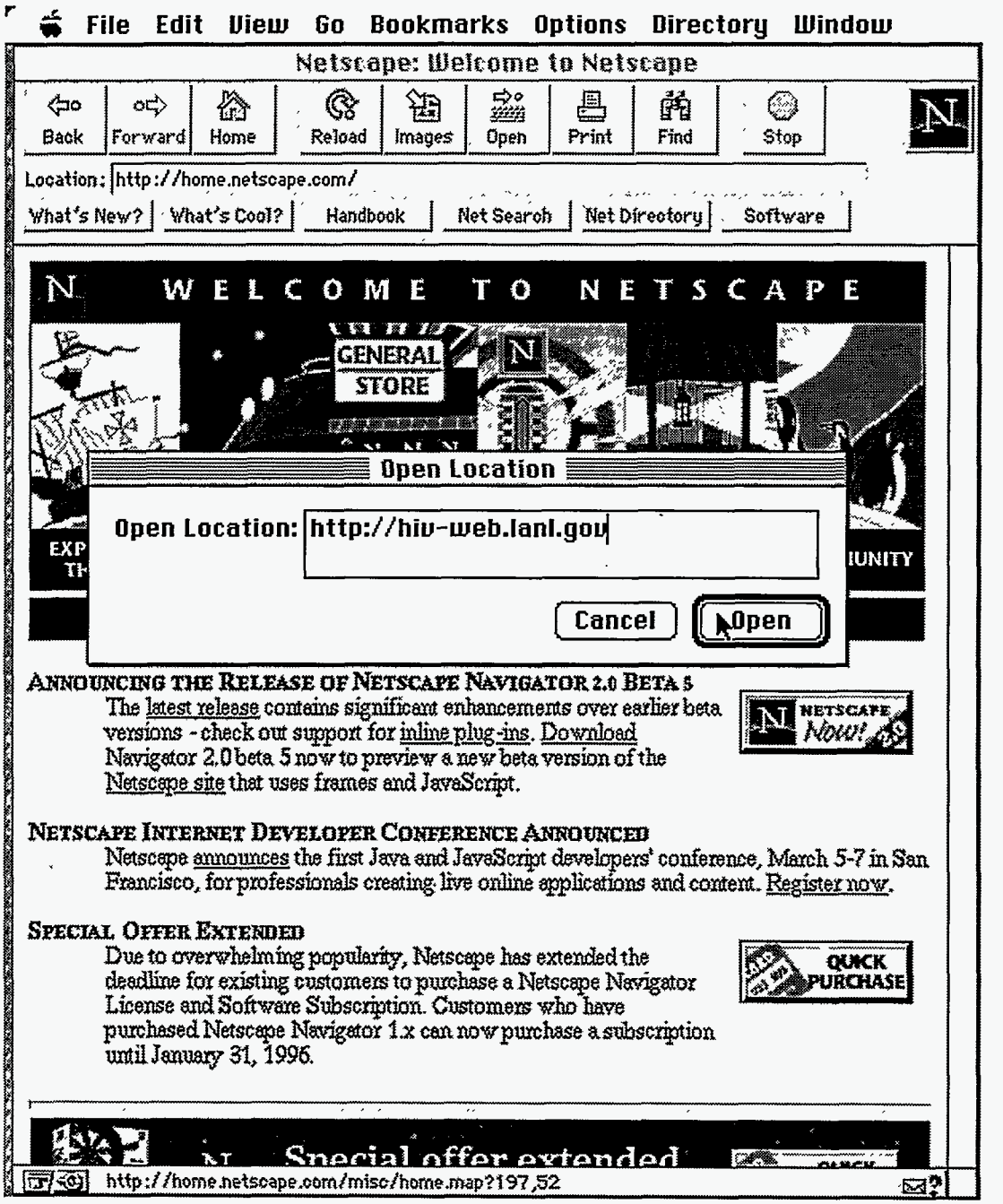
Tour of the World Wide Web

We reprint here the short tour of our Web site which we originally published in last year's Compendium. It will give the uninitiated a feel for how the Web works. You begin by "launching" your browser program. (The browser in the illustrations here is Netscape.) In the Netscape window on your screen there appears the starting point for your navigation of the Web (Screen 1). This starting point can be any address you have specified, it could even be the HIV database URL; by default it is the Netscape "home page."



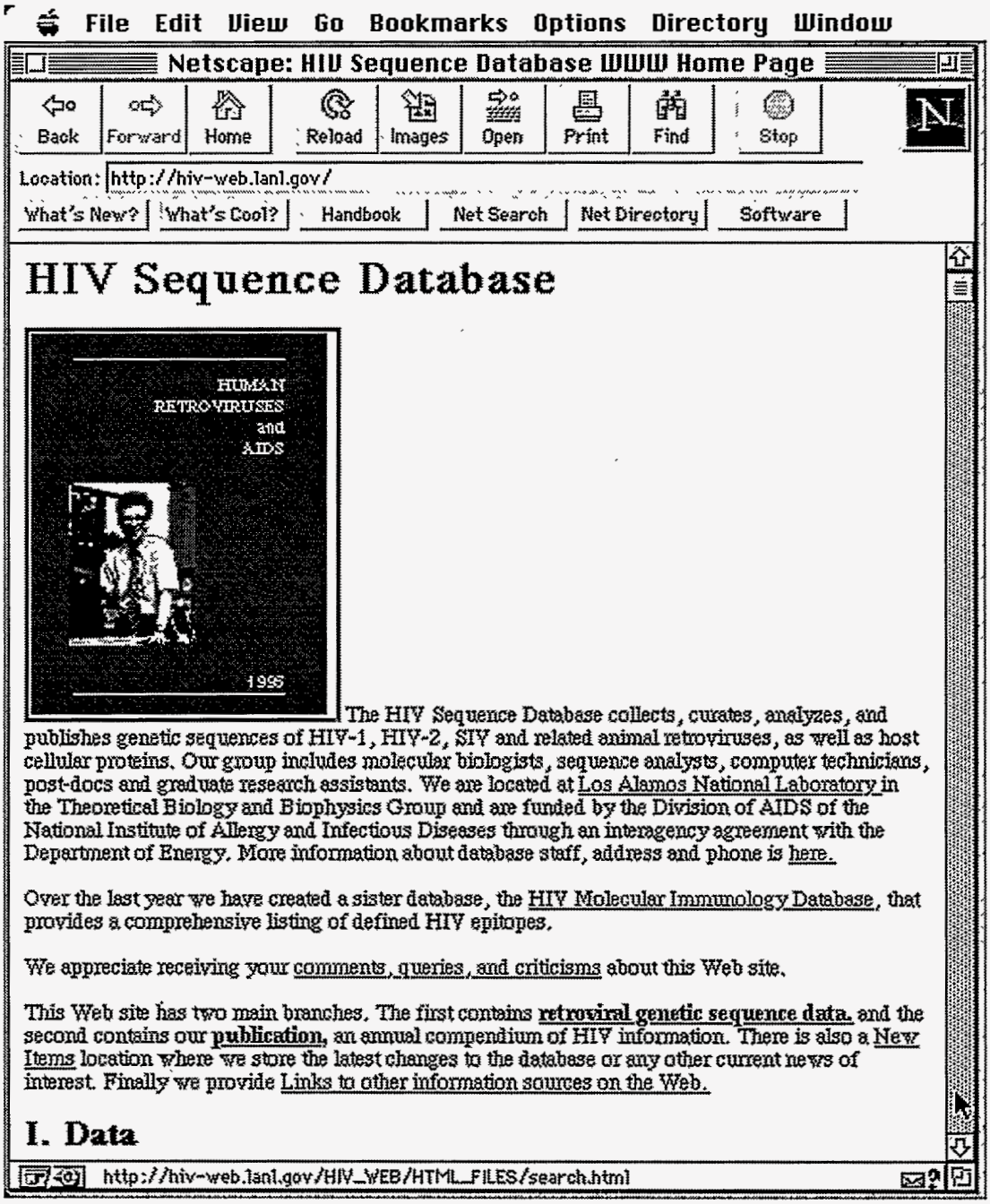
Screen 1

From your starting point you select "Open Location" from the File menu in the upper left corner of the window. A small window entitled Open Location appears, you type in the URL of the HIV Database `http://hiv-web.lanl.gov` in the space provided (Screen 2), and click Open. You can also store frequently visited URLs in a list called "bookmarks" or "hotlist" and select your destination from this list rather than typing its URL.



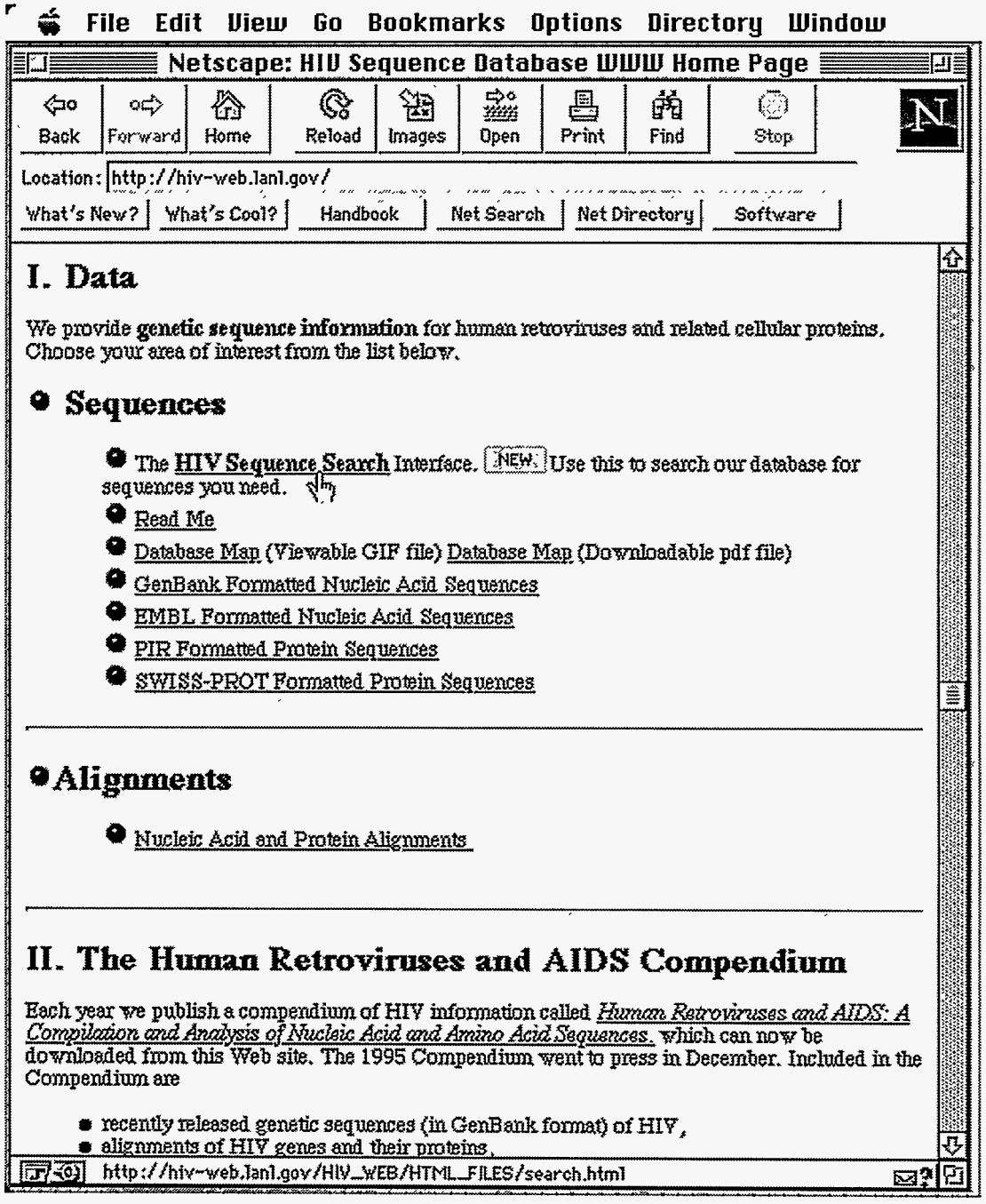
Screen 2

After a few seconds a new image appears in the window (Screen 3), the home page of the HIV Database. The text in the window is formatted and pleasantly readable. All the text on the home page cannot fit on the screen at once, but you can scroll through it by clicking your mouse on the down arrow in the lower right.



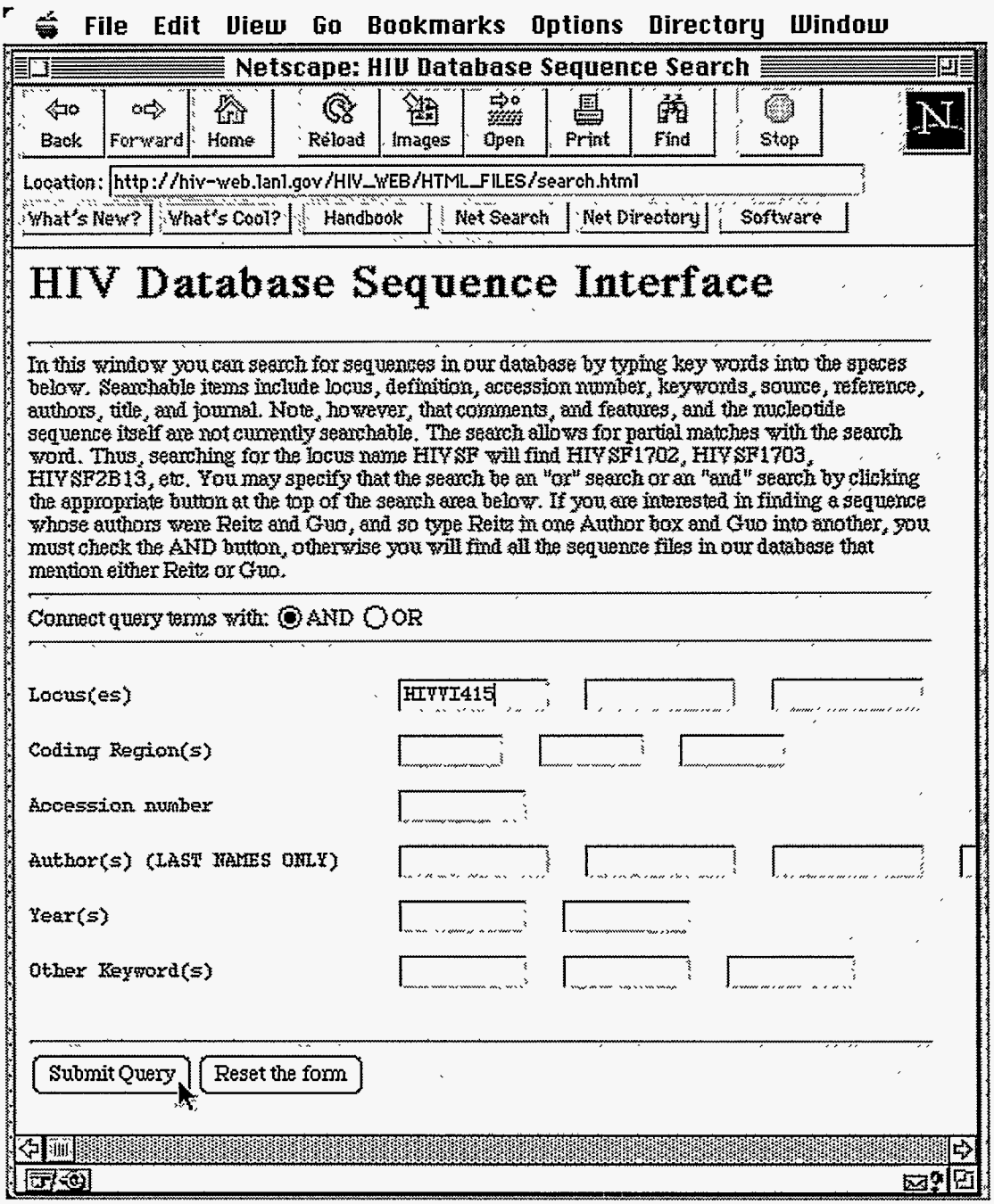
Screen 3

As you move down in the document you come to an area labeled "I. Data" (Screen 4). Some of the text is underlined and, on color monitors, shows in a different color. These underlined words are links to other locations on the WWW. The locations may be local or they may be on computers half way around the world; the links to them look exactly the same and are activated in the same way. If we position our cursor arrow at one of the links, as we have in Screen 4, the arrow turns into a pointing hand. Pressing the mouse button transfers us quickly and automatically to the location attached to that link. In our example the finger points to HIV Sequence Search.



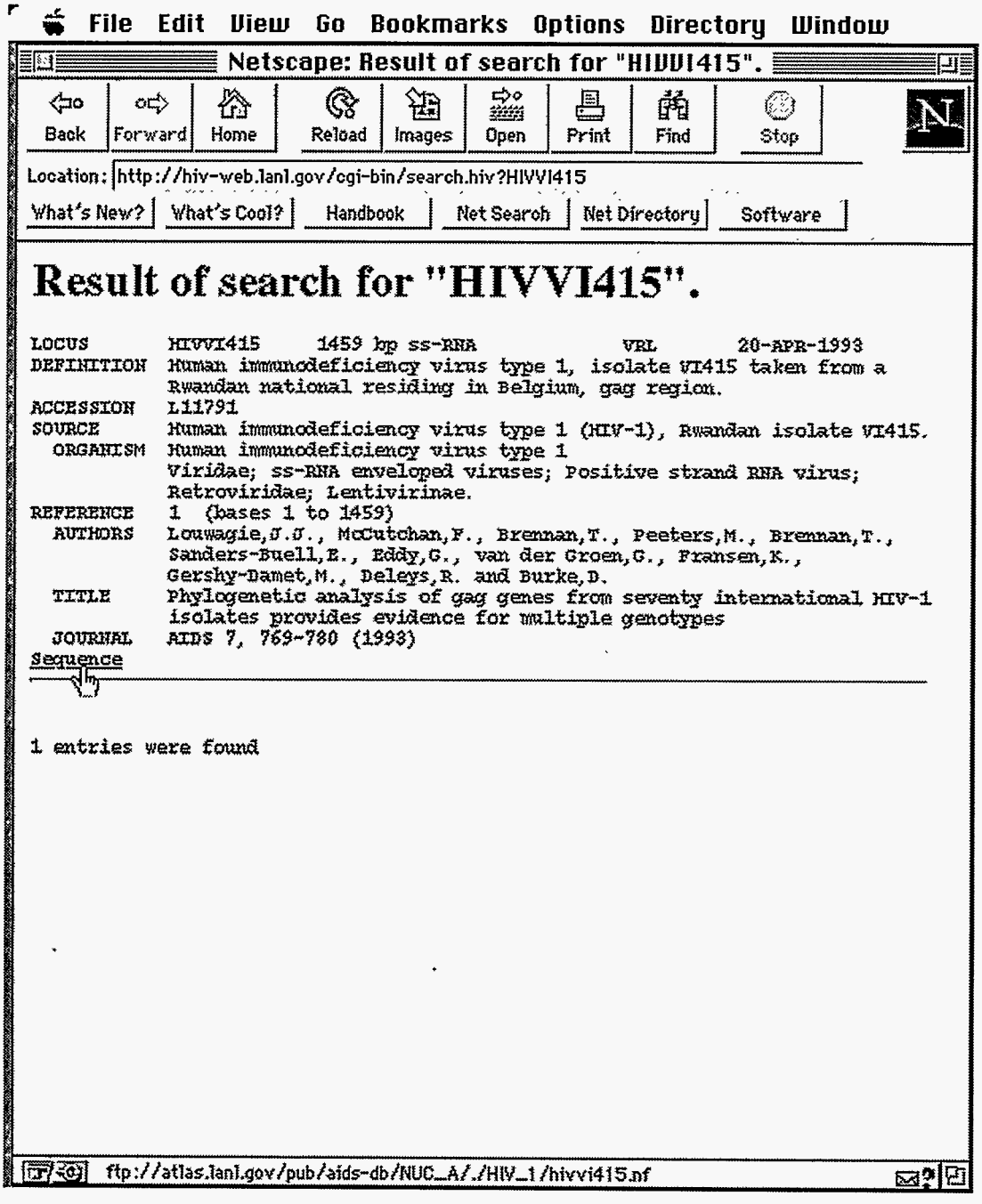
Screen 4

Upon clicking HIV Sequence Search the search area appears on the screen (Screen 5). An explanation of how to use the search utility is included. We type in a search pattern—in this example HIVVI415, a GenBank locus name—and hit the carriage return.



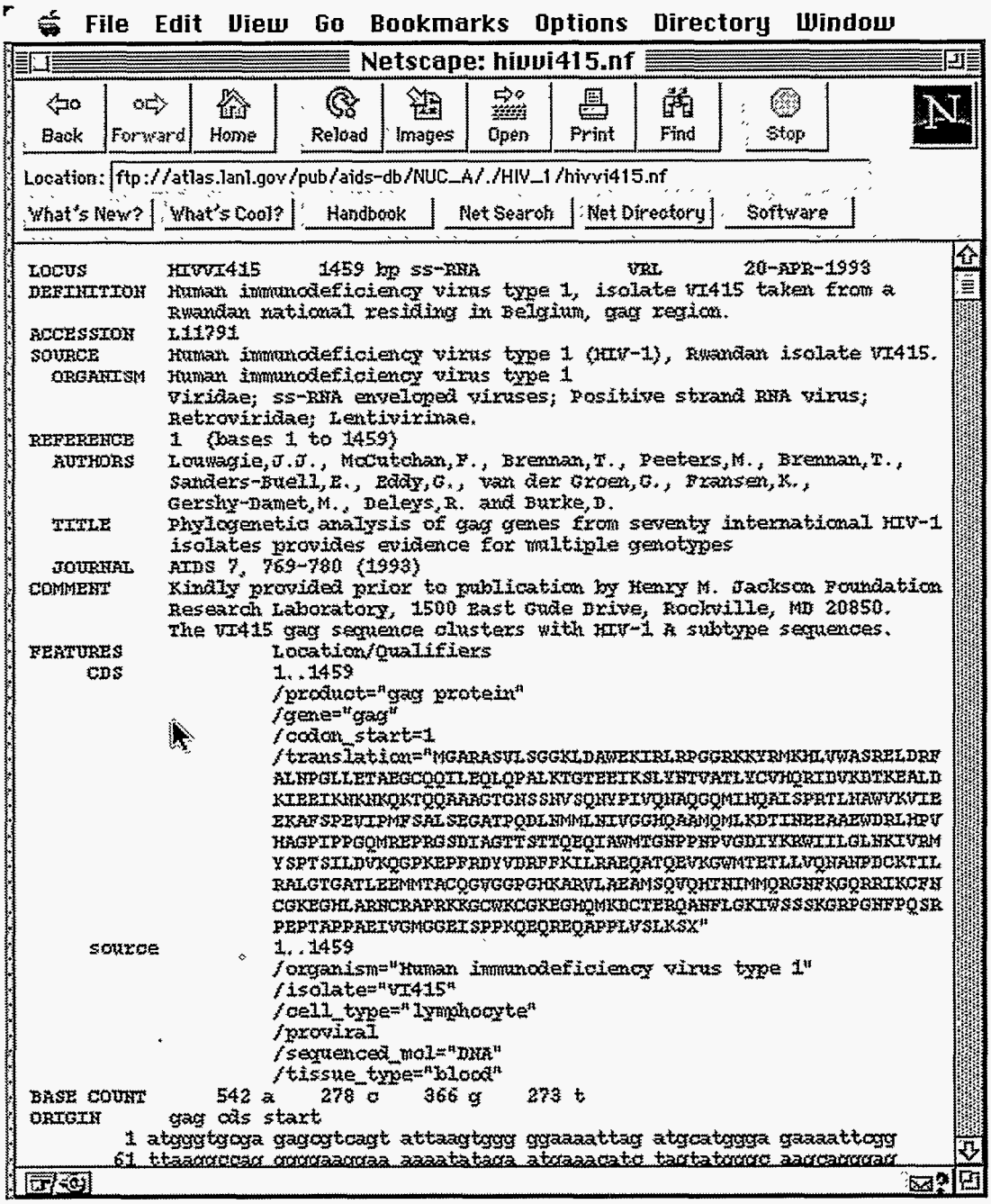
Screen 5

The result of the search, an abbreviated GenBank file, appears (Screen 6). We can retrieve the entire file by positioning the pointing hand on the link in Screen 6 called Sequence and clicking.



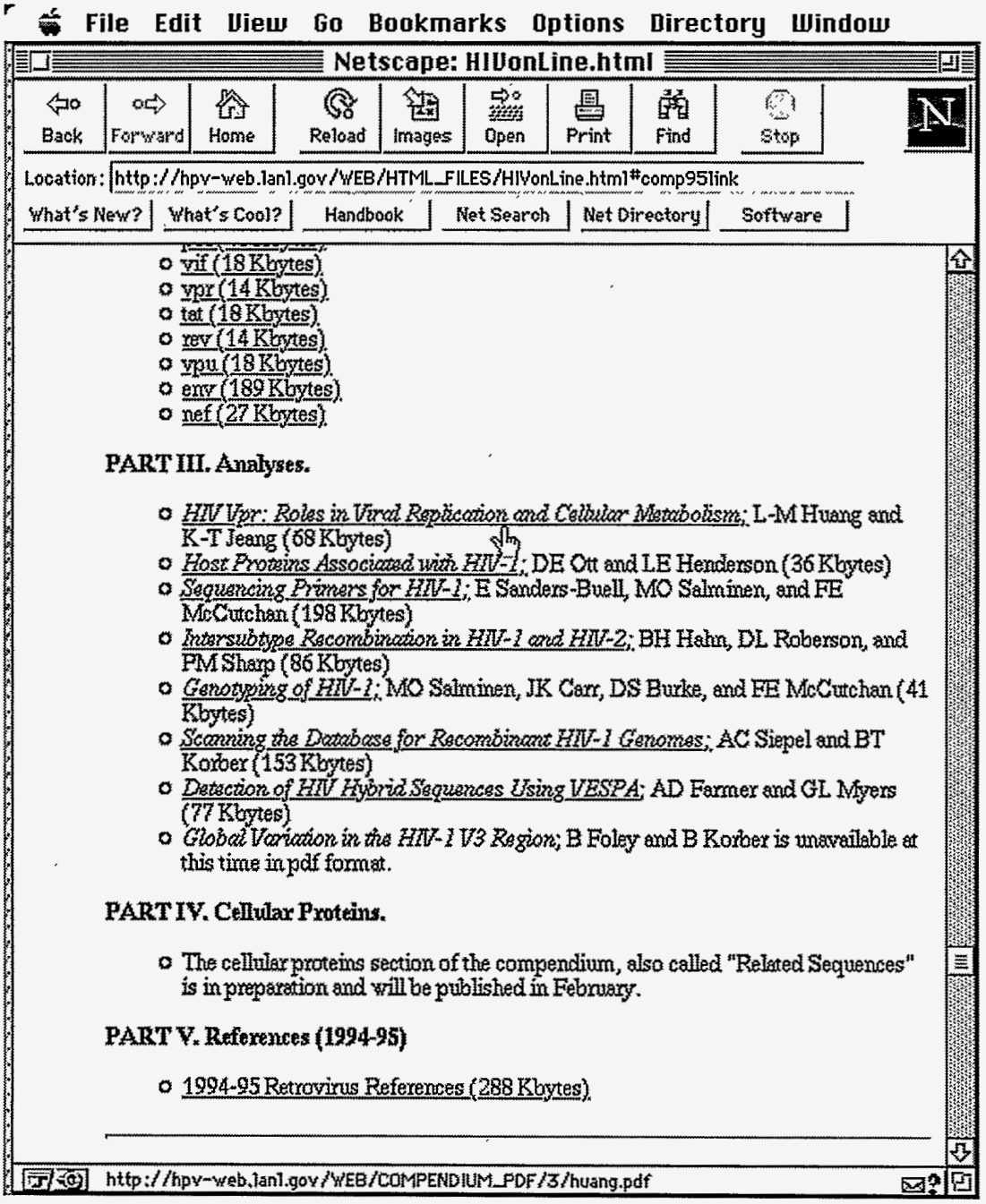
Screen 6

The result is shown in Screen 7, the complete GenBank file. This file may be transferred to your local computer in its entirety by clicking on "File" in the menu bar at the top of the screen and selecting "Save As . . . " from that menu.



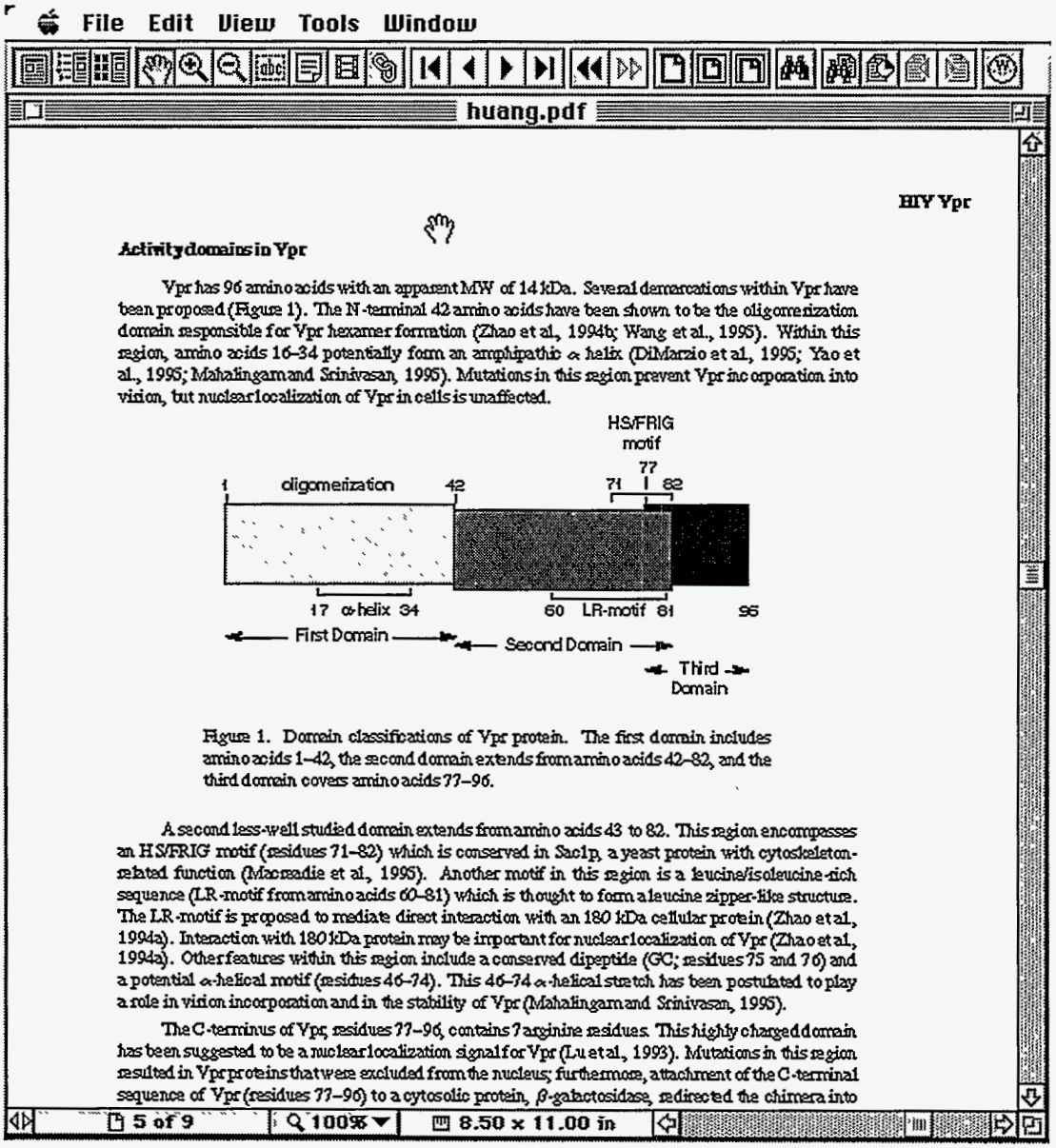
Screen 7

As we follow a series of links on the WWW, the browser remembers the places we have been and allows us to return to them by clicking the "Back" button located at the top left corner of the window. If we return to the HIV home page (Screen 3) we can follow a different path of links (not shown here) that carries us to another area of the site, the "on-line version of the *Human Retroviruses and AIDS* compendium." Screen 8 shows a part of the table of contents. The pointing hand is about to retrieve an article on HIV Vpr. This link, though it looks the same as others, behaves differently. A copy of the Vpr article is automatically transferred to your local computer and the Adobe Acrobat viewer is automatically launched. Acrobat is a "helper application" used to read files that are in a format which is nonstandard for the WWW. Our Web site shows you how to obtain a copy of the Acrobat viewer.



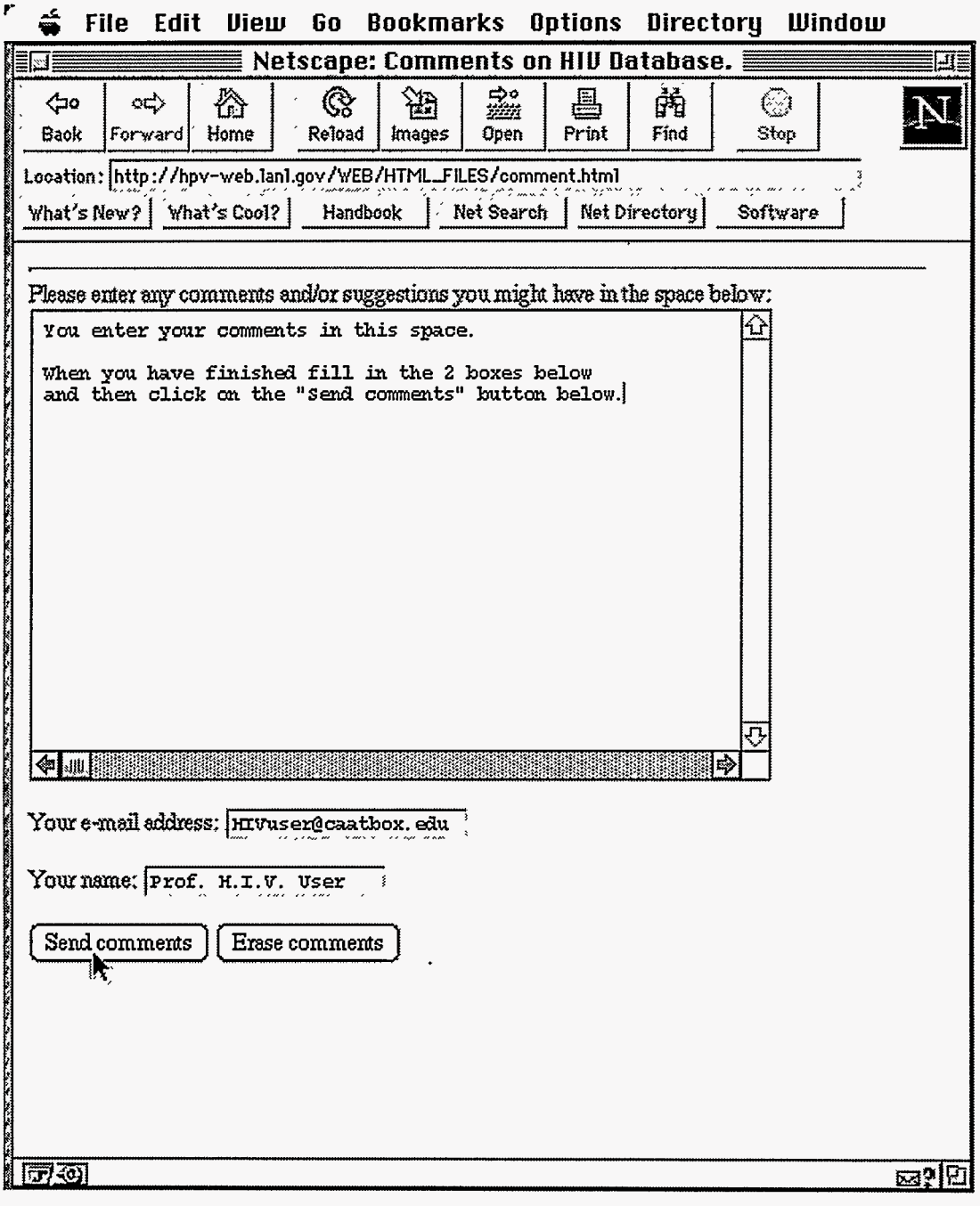
Screen 8

The Acrobat viewer, available free from the WWW (`http://www.adobe.com/Acrobat`), displays the Vpr article (Screen 9), which is written in "portable document format" (PDF) and allows you to print out a perfect copy of it, exactly as it appears in the printed compendium. Acrobat viewer has a search mechanism that allows you to search in a document for words that might lead you to a subject of interest. For example, Part V of the compendium lists hundreds of HIV references. Those references are searchable with Acrobat Viewer. Searches for specific nucleotide sequences in Part I (e.g. `AATAAA`) will be less useful because of the way GenBank files are formatted, with spaces and sequence numbers embedded in the file.



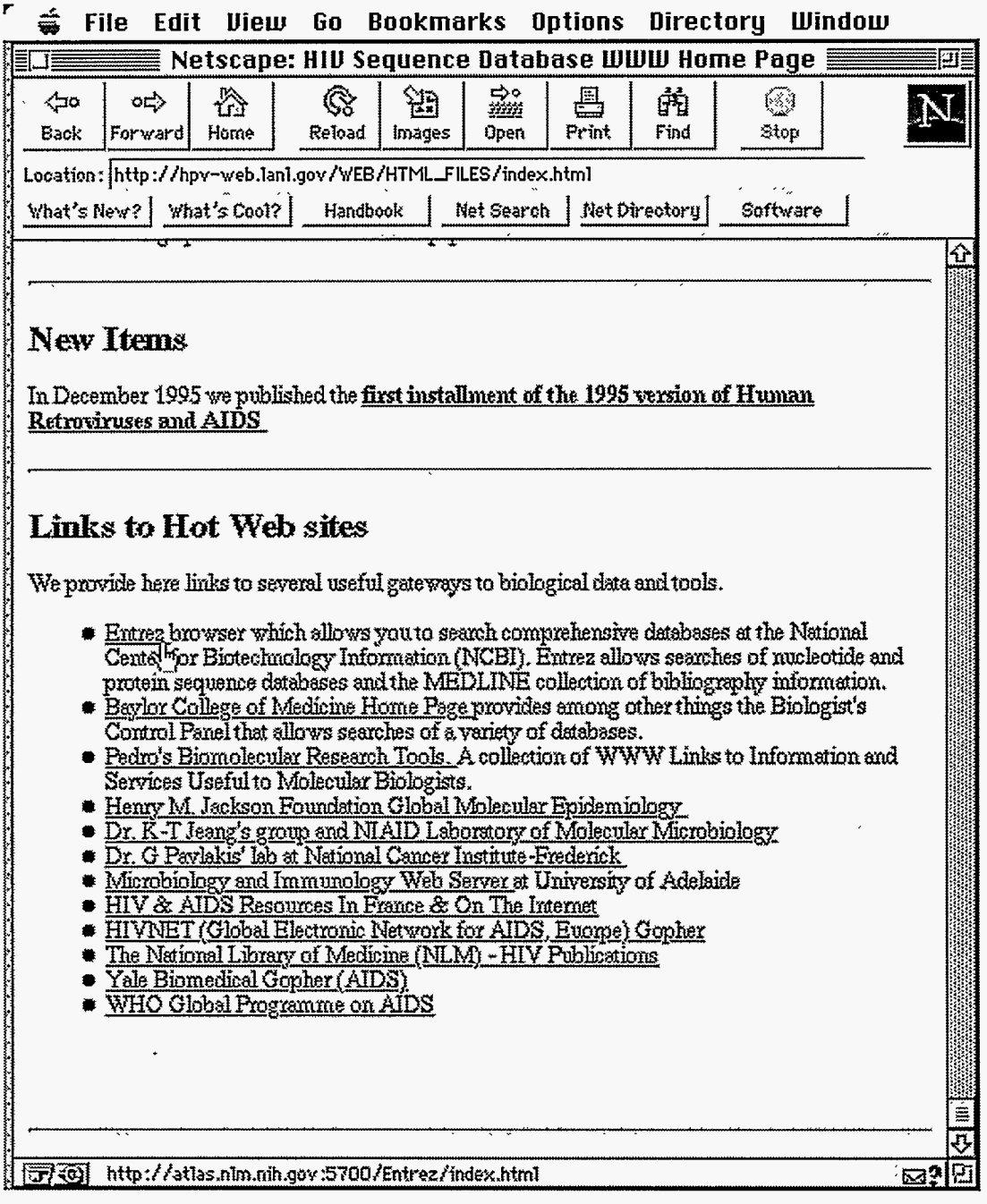
Screen 9

Our Web site provides users an opportunity to communicate with us from several points on the site. An electronic "comments" form (Screen 10) can be activated that allows you to write us a letter to which we promise to respond.



Screen 10

Finally, we have connected our site to several other locations of utility to biologists as shown on Screen 11, below.



Screen 11

Data Access by FTP

The complete database of sequences and alignments are also available electronically from a server here at Los Alamos called "atlas". In the first section of Part V we described accessing this, and other, information using the WWW. You may also retrieve information from atlas by the "anonymous ftp" method. In the following instructions for accessing atlas, what you actually type at your terminal is shown in **boldface** type [with an explanation shown in square brackets]. What the server responds with is shown indented in italic type.

---

ftp atlas.lanl.gov [this ftp command connects your computer to atlas]
Name (atlas.lanl.gov): [atlas asks for your user name]
anonymous [for offsite users anonymous is the only name that works]
Guest login ok, send ident as password.
Password:
[enter your e-mail address or your name as ident]
Guest login ok. access restrictions apply.
cd pub [change to the directory called pub]
CWD command successful
cd aids-db
ls [this lists the files and directories inside of aids-db]
AIDS_A.TAR.Z, AIDS_E.TAR.Z, NUC_A, PRO_A, ALIGN, NUC_E, PRO_E, PROGS, README.TXT, ROADMAP
pwd [to show your location in the file hierarchy]
"/pub/aids-db" is current directory.
cd NUC_A/HIV_1 [move down 2 directories]
ls [all the GenBank format files are listed]
get hivr1s4 [this copies the file hivr1s4 to your home computer]
mget *.nf [this copies all files in the current directory with the suffix .nf to your home computer]
bye [this disconnects you from the server]
A map of the aids-db directories and files on atlas is reproduced on the following page. Many of the directories contain ReadMe files that describe the contents.

Aberham C; Weber S; Phares W
Spontaneous mutations in the human immunodeficiency virus type 1 gag gene that affect viral replication in the presence of cyclosporins.
J Virol 70: 3536-44 (1996)

Adleman LM; Wofsy D
Blind T-cell homeostasis in CD4-deficient mice.
J Acquir Immune Defic Syndr Hum Retrovirol 11: 334-40 (1996)

Agy MB; Sherbert CH; Katze MG
Development of an in vitro mRNA degradation assay utilizing extracts from HIV-1- and SIV-infected cells.
Virology 217: 158-66 (1996)

Agy MB; Sherbert CH; Katze MG
Development of an in vitro mRNA degradation assay utilizing extracts from HIV-1- and SIV-infected cells.
Virology 217: 158-66 (1996)

Aiken C; Krause L; Chen YL; Trono D
Mutational analysis of HIV-1 Nef: identification of two mutants that are temperature-sensitive for CD4 downregulation.
Virology 217: 293-300 (1996)

Akari H; Mori K; Terao K; Otani I; Fukasawa M; Mukai R; Yoshikawa Y
In vitro immortalization of Old World monkey T lymphocytes with Herpesvirus saimiri: its susceptibility to infection with simian immunodeficiency viruses.
Virology 218: 382-8 (1996)

Akashi A; Ono S; Kuwano K; Arai S
Proteins of 30 and 36 kilodaltons, membrane constituents of the Staphylococcus aureus L form, induce production of tumor necrosis factor alpha and activate the human immunodeficiency virus type 1 long terminal repeat.
Infect Immun 64: 3267-72 (1996)

Akridge RE; Reed SG
Interleukin-12 decreases human immunodeficiency virus type 1 replication in human macrophage cultures reconstituted with autologous peripheral blood mononuclear cells.
J Infect Dis 173: 559-64 (1996)

Albini A; Benelli R; Presta M; Rusnati M; Ziche M; Rubartelli A; Paglialunga G; Bussolino F; Noonan D
HIV-tat protein is a heparin-binding angiogenic growth factor.
Oncogene 12: 289-97 (1996)

Aldrovandi GM; Zack JA
Replication and pathogenicity of human immunodeficiency virus type 1 accessory gene mutants in SCID-hu mice.
J Virol 70: 1505-11 (1996)

Alin K; Goff SP
Mutational analysis of interactions between the Gag precursor proteins of murine leukemia viruses.
Virology 216: 418-24 (1996)

Althaus IW; Franks KM; Langley KB; Kezdy FJ; Peterson T; Buxser SE; Decker DE; Dolak LA; Ulrich RG; Reusser F
The amphiphilic properties of novenamines determine their activity as inhibitors of HIV-1 RNase H.
Experientia 52: 329-35 (1996)

Altman JD; Moss PAH; Goulder PJR; Barouch DH; McHeyzer-Williams MG; Bell JI; McMichael AJ; Davis MM
Phenotypic analysis of antigen-specific T lymphocytes.
Science 274: 94-6 (1996)

Altman JD; Moss PAH; Goulder PJR; Barouch DH; McHeyzer-Williams MG; Bell JI; McMichael AJ; Davis MM
Phenotypic analysis of antigen-specific T lymphocytes.
Science 274: 94-6 (1996)

Amakawa R; Hakem A; Kundig TM; Matsuyama T; Simard JJ; Timms E; Wakeham A; Mittruecker HW; Griesser H; Takimoto H; Schmits R; Shahinian A; Ohashi P; Penninger JM; Mak TW
Impaired negative selection of T cells in Hodgkin's disease antigen CD30-deficient mice.
Cell 84: 551-62 (1996)

Anderson MJ; Porter DC; Fultz PN; Morrow CD
Poliovirus replicons that express the gag or the envelope surface protein of simian immunodeficiency virus SIV(smm) PBj14.
Virology 219: 140-9 (1996)

Andres JM
Neonatal hepatobiliary disorders.
Clin Perinatol 23: 321-52 (1996)

Andres JM
Neonatal hepatobiliary disorders.
Clin Perinatol 23: 321-52 (1996)

Ankel H; Capobianchi MR; Frezza F; Castilletti C; Dianzani F
Interferon induction by HIV-1-infected cells: a possible role of sulfatides or related glycolipids.
Virology 221: 113-9 (1996)

Ansari-Lari MA; Gibbs RA
Expression of human immunodeficiency virus type 1 reverse transcriptase in trans during virion release and after infection.
J Virol 70: 3870-5 (1996)

Arase N; Arase H; Ohki K; Nishino Y; Ikuta K; Onoe K
Mitogenic effect of HIV-infected human T cell lines on mouse B cells mediated by surface immunoglobulin.
Clin Exp Immunol 103: 24-9 (1996)

Arion D; Borkow G; Gu Z; Wainberg MA; Parniak MA
The K65R mutation confers increased DNA polymerase processivity to HIV-1 reverse transcriptase.
J Biol Chem 271: 19860-4 (1996)

Arion D; Borkow G; Gu Z; Wainberg MA; Parniak MA
The K65R mutation confers increased DNA polymerase processivity to HIV-1 reverse transcriptase.
J Biol Chem 271: 19860-4 (1996)

Arion D; Harada R; Li X; Wainberg MA; Parniak MA
HIV-1 reverse transcriptase shows no specificity for the binding of primer tRNA(Lys3).
Biochem Biophys Res Commun 225: 839-43 (1996)

Arion D; Harada R; Li X; Wainberg MA; Parniak MA
HIV-1 reverse transcriptase shows no specificity for the binding of primer tRNA(Lys3).
Biochem Biophys Res Commun 225: 839-43 (1996)

Artenstein AW; Hegerich PA; Beyrer C; Rungruengthanakit K; Michael NL; Natpratan C
Sequences and phylogenetic analysis of the nef gene from Thai subjects harboring subtype E HIV-1.
AIDS Res Hum Retroviruses 12: 557-60 (1996)

Arts EJ; Ghosh M; Jacques PS; Ehresmann B; Le Grice SF
Restoration of tRNA3Lys-primed(-)-strand DNA synthesis to an HIV-1 reverse transcriptase mutant with extended tRNAs. Implications for retroviral replication.
J Biol Chem 271: 9054-61 (1996)

Arts EJ; Marois JP; Gu Z; Le Grice SF; Wainberg MA
Effects of 3'-deoxynucleoside 5'-triphosphate concentrations on chain termination by nucleoside analogs during human immunodeficiency virus type 1 reverse transcription of minus-strand strong-stop DNA.
J Virol 70: 712-20 (1996)

Artzi HB; Shemesh J; Zeelon E; Amit B; Kleiman L; Gorecki M; Panet A
Ribonuclease H activity during initiation of reverse transcription using tRNA(lys)/RNA primer/template of human immunodeficiency virus.
Arch Biochem Biophys 325: 209-16 (1996)

Aruoma OI; Spencer JP; Rossi R; Aeschbach R; Khan A; Mahmood N; Munoz A; Murcia A; Butler J; Halliwell B
An evaluation of the antioxidant and antiviral action of extracts of rosemary and Provencal herbs.
Food Chem Toxicol 34: 449-56 (1996)

Arya SK; Gallo RC
Human immunodeficiency virus (HIV) type 2-mediated inhibition of HIV type 1: a new approach to gene therapy of HIV-infection.
Proc Natl Acad Sci U S A 93: 4486-91 (1996)

Arya SK; Gallo RC
Human immunodeficiency virus (HIV) type 2-mediated inhibition of HIV type 1: a new approach to gene therapy of HIV-infection.
Proc Natl Acad Sci U S A 93: 4486-91 (1996)

Asante-Appiah E; Chan WW
Synergistic binding of inhibitors to the protease from HIV type 1.
Biochem J 315 (Pt 1): 113-7 (1996)

Asch S; Knowles L; Rai A; Jones BE; Pogoda J; Barnes PF
Relationship of isoniazid resistance to human immunodeficiency virus infection in patients with tuberculosis [see comments]
Am J Respir Crit Care Med 153: 1708-10 (1996)

Askari FK; McDonnell WM
Antisense-oligonucleotide therapy.
N Engl J Med 334: 316-8 (1996)

Atkins MC; Carlin EM; Emery VC; Griffiths PD; Boag F
Fluctuations of HIV load in semen of HIV positive patients with newly acquired sexually transmitted diseases.
BMJ 313: 341-2 (1996)

Avivi I; Blumenfeld I; Blumenfeld Z
[Therapeutic potential of ozone–possible side-effects]
Harefuah 130: 327-30 (1996)

Ayyavoo V; Ugen KE; Fernandes LS; Goedert JJ; Rubinstein A; Williams WV; Weiner DB
Analysis of genetic heterogeneity, antigenicity, and biological characteristics of HIV-1 in a maternal transmitter and nontransmitter patient pair.
DNA Cell Biol 15: 571-80 (1996)

Back NK; Nijhuis M; Keulen W; Boucher CA; Oude Essink BO; van Kuilenburg AB; van Gennip AH; Berkhout B
Reduced replication of 3TC-resistant HIV-1 variants in primary cells due to a processivity defect of the reverse transcriptase enzyme.
EMBO J 15: 4040-9 (1996)

Back NK; Nijhuis M; Keulen W; Boucher CA; Oude Essink BO; van Kuilenburg AB; van Gennip AH; Berkhout B
Reduced replication of 3TC-resistant HIV-1 variants in primary cells due to a processivity defect of the reverse transcriptase enzyme.
EMBO J 15: 4040-9 (1996)

Badley AD; McElhinny JA; Leibson PJ; Lynch DH; Alderson MR; Paya CV
Upregulation of Fas ligand expression by human immunodeficiency virus in human macrophages mediates apoptosis of uninfected T lymphocytes.
J Virol 70: 199-206 (1996)

Bagetta G; Corasaniti MT; Aloe L; Berliocchi L; Costa N; Finazzi-Agro A; Nistico G
Intracerebral injection of human immunodeficiency virus type 1 coat protein gp120 differentially affects the expression of nerve growth factor and nitric oxide synthase in the hippocampus of rat.
Proc Natl Acad Sci U S A 93: 928-33 (1996)

Bahner I; Kearns K; Hao QL; Smogorzewska EM; Kohn DB
Transduction of human CD34+ hematopoietic progenitor cells by a retroviral vector expressing an RRE decoy inhibits human immunodeficiency virus type 1 replication in myelomonocytic cells produced in long-term culture.
J Virol 70: 4352-60 (1996)

Bailly C; Colson P; Houssier C; Hamy F
The binding mode of drugs to the TAR RNA of HIV-1 studied by electric linear dichroism.
Nucleic Acids Res 24: 1460-4 (1996)

Bakhanashvili M; Avidan O; Hizi A
Mutational studies of human immunodeficiency virus type 1 reverse transcriptase: the involvement of residues 183 and 184 in the fidelity of DNA synthesis.
FEBS Lett 391: 257-62 (1996)

Balachandran R; Singh MK; Gupta P
An additional mechanism of growth restriction in T cell line H9 of human immunodeficiency virus type 1 isolates from asymptomatic homosexual men.
J Gen Virol 77 (Pt 5): 1083-8 (1996)

Balakrishnan M; Zastrow D; Jonsson CB
Catalytic activities of the human T-cell leukemia virus type II integrase.
Virology 219: 77-86 (1996)

Balzarini J; Aquaro S; Perno CF; Witvrouw M; Holy A; De Clercq E
Activity of the (R)-enantiomers of 9-(2-phosphonylmethoxypropyl)-adenine and 9-(2-phosphonylmethoxypropyl)-2,6-diaminopurine against human immunodeficiency virus in different human cell systems.
Biochem Biophys Res Commun 219: 337-41 (1996)

Balzarini J; Pelemans H; Aquaro S; Perno CF; Witvrouw M; Schols D; De Clercq E; Karlsson A
Highly favorable antiviral activity and resistance profile of the novel thiocarboxanilide pentenyloxy ether derivatives UC-781 and UC-82 as inhibitors of human immunodeficiency virus type 1 replication.
Mol Pharmacol 50: 394-401 (1996)

Balzarini J; Pelemans H; Perez-Perez MJ; San-Felix A; Camarasa MJ; De Clercq E; Karlsson A
Marked inhibitory activity of non-nucleoside reverse transcriptase inhibitors against human immunodeficiency virus type 1 when combined with (-)2',3'-dideoxy-3'-thiacytidine.
Mol Pharmacol 49: 882-90 (1996)

Balzarini J; Wedgwood O; Kruining J; Pelemans H; Heijtink R; De Clercq E; McGuigan C
Anti-HIV and anti-HBV activity and resistance profile of 2',3'-dideoxy-3'-thiacytidine (3TC) and its arylphosphoramidate derivative CF 1109.
Biochem Biophys Res Commun 225: 363-9 (1996)

Balzarini J; Wedgwood O; Kruining J; Pelemans H; Heijtink R; De Clercq E; McGuigan C
Anti-HIV and anti-HBV activity and resistance profile of 2',3'-dideoxy-3'-thiacytidine (3TC) and its arylphosphoramidate derivative CF 1109.
Biochem Biophys Res Commun 225: 363-9 (1996)

Barnes PF; el-Hajj H; Preston-Martin S; Cave MD; Jones BE; Otaya M; Pogoda J; Eisenach KD
Transmission of tuberculosis among the urban homeless.
JAMA 275: 305-7 (1996)

Barnum D; Greene J; Smellie A; Sprague P
Identification of common functional configurations among molecules.
J Chem Inf Comput Sci 36: 563-71 (1996)

Barratt-Boyes SM; Henderson RA; Finn OJ
Chimpanzee dendritic cells with potent immunostimulatory function can be propagated from peripheral blood.
Immunology 87: 528-34 (1996)

Barre-Sinoussi F
HIV as the cause of AIDS.
Lancet 348: 31-5 (1996)

Barrie KA; Perez EE; Lamers SL; Farmerie WG; Dunn BM; Sleasman JW; Goodenow MM
Natural variation in HIV-1 protease, Gag p7 and p6, and protease cleavage sites within gag/pol polyproteins: amino acid substitutions in the absence of protease inhibitors in mothers and children infected by human immunodeficiency virus type 1.
Virology 219: 407-16 (1996)

Barsov EV; Huber WE; Marcotrigiano J; Clark PK; Clark AD; Arnold E; Hughes SH
Inhibition of human immunodeficiency virus type 1 integrase by the Fab fragment of a specific monoclonal antibody suggests that different multimerization states are required for different enzymatic functions.
J Virol 70: 4484-94 (1996)

Barton CH; Biggs TE; Mee TR; Mann DA
The human immunodeficiency virus type 1 regulatory protein Tat inhibits interferon-induced iNos activity in a murine macrophage cell line.
J Gen Virol 77 (Pt 8): 1643-7 (1996)

Bartz SR; Rogel ME; Emerman M
Human immunodeficiency virus type 1 cell cycle control: Vpr is cytostatic and mediates G2 accumulation by a mechanism which differs from DNA damage checkpoint control.
J Virol 70: 2324-31 (1996)

Battiste JL; Mao H; Rao NS; Tan R; Muhandiram DR; Kay LE; Frankel AD; Williamson JR
Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex.
Science 273: 1547-51 (1996)

Battiste JL; Mao H; Rao NS; Tan R; Muhandiram DR; Kay LE; Frankel AD; Williamson JR
Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex.
Science 273: 1547-51 (1996)

Baumler CB; Bohler T; Herr I; Benner A; Krammer PH; Debatin KM
Activation of the CD95 (APO-1/Fas) system in T cells from human immunodeficiency virus type-1-infected children.
Blood 88: 1741-6 (1996)

Baumler CB; Bohler T; Herr I; Benner A; Krammer PH; Debatin KM
Activation of the CD95 (APO-1/Fas) system in T cells from human immunodeficiency virus type-1-infected children.
Blood 88: 1741-6 (1996)

Beale J
Reconsidering an attenuated vaccine for AIDS.
Lancet 347: 344-5 (1996)

Beard WA; Minnick DT; Wade CL; Prasad R; Won RL; Kumar A; Kunkel TA; Wilson SH
Role of the "helix clamp" in HIV-1 reverse transcriptase catalytic cycling as revealed by alanine-scanning mutagenesis.
J Biol Chem 271: 12213-20 (1996)

Beatty JA; Willett BJ; Gault EA; Jarrett O
A longitudinal study of feline immunodeficiency virus-specific cytotoxic T lymphocytes in experimentally infected cats, using antigen-specific induction.
J Virol 70: 6199-206 (1996)

Beauparlant P; Kwon H; Clarke M; Lin R; Sonenberg N; Wainberg M; Hiscott J
Transdominant mutants of I kappa B alpha block Tat-tumor necrosis factor synergistic activation of human immunodeficiency virus type 1 gene expression and virus multiplication.
J Virol 70: 5777-85 (1996)

Beer B; Scherer J; zur Megede J; Norley S; Baier M; Kurth R
Lack of dichotomy between virus load of peripheral blood and lymph nodes during long-term simian immunodeficiency virus infection of African green monkeys.
Virology 219: 367-75 (1996)

Beer B; Scherer J; zur Megede J; Norley S; Baier M; Kurth R
Lack of dichotomy between virus load of peripheral blood and lymph nodes during long-term simian immunodeficiency virus infection of African green monkeys.
Virology 219: 367-75 (1996)

Beissinger M; Paulus C; Bayer P; Wolf H; Rosch P; Wagner R
Sequence-specific resonance assignments of the 1H-NMR spectra and structural characterization in solution of the HIV-1 transframe protein p6.
Eur J Biochem 237: 383-92 (1996)

Belzhelarskaia SN; Sutugina LP; Tolkacheva TV; Rubtsov PM; Skriabin KG
[Expression of the HIV-1 CD4 receptor gene in insect cells using a baculoviral system]
Mol Biol (Mosk) 30: 524-8 (1996)

Belzhelarskaia SN; Sutugina LP; Tolkacheva TV; Rubtsov PM; Skriabin KG
[Expression of the HIV-1 CD4 receptor gene in insect cells using a baculoviral system]
Mol Biol (Mosk) 30: 524-8 (1996)

Ben-Artzi H; Shemesh J; Zeelon E; Amit B; Kleiman L; Gorecki M; Panet A
Molecular analysis of the second template switch during reverse transcription of the HIV RNA template.
Biochemistry 35: 10549-57 (1996)

Ben-Artzi H; Shemesh J; Zeelon E; Amit B; Kleiman L; Gorecki M; Panet A
Molecular analysis of the second template switch during reverse transcription of the HIV RNA template.
Biochemistry 35: 10549-57 (1996)

Benjamin D; Sharma V; Kubin M; Klein JL; Sartori A; Holliday J; Trinchieri G
IL-12 expression in AIDS-related lymphoma B cell lines.
J Immunol 156: 1626-37 (1996)

Benveniste O; Vaslin B; Le Grand R; Cheret A; Matheux F; Theodoro F; Cranage MP; Dormont D
Comparative interleukin (IL-2)/interferon IFN-gamma and IL-4/IL-10 responses during acute infection of macaques inoculated with attenuated nef-truncated or pathogenic SIC-mac251 virus.
Proc Natl Acad Sci U S A 93: 3658-63 (1996)

Berkhout B; van Wamel JL
Identification of a novel splice acceptor in the HIV-1 genome: independent expression of the cytoplasmic tail of the envelope protein.
Arch Virol 141: 839-55 (1996)

Berman PW; Murthy KK; Wrin T; Vennari JC; Cobb EK; Eastman DJ; Champe M; Nakamura GR; Davison D; Powell MF; Bussiere J; Francis DP; Matthews T; Gregory TJ; Obijeski JF
Protection of MN-rgp120-immunized chimpanzees from heterologous infection with a primary isolate of human immunodeficiency virus type 1.
J Infect Dis 173: 52-9 (1996)

Berson JF; Long D; Doranz BJ; Rucker J; Jirik FR; Doms RW
A seven-transmembrane domain receptor involved in fusion and entry of T-cell-tropic human immunodeficiency virus type 1 strains.
J Virol 70: 6288-95 (1996)

Berthold E; Maldarelli F
cis-acting elements in human immunodeficiency virus type 1 RNAs direct viral transcripts to distinct intranuclear locations.
J Virol 70: 4667-82 (1996)

Berube P; Barbeau B; Cantin R; Sekaly RP; Tremblay M
Repression of human immunodeficiency virus type 1 long terminal repeat-driven gene expression by binding of the virus to its primary cellular receptor, the CD4 molecule.
J Virol 70: 4009-16 (1996)

Bessudo A; Cherepakhin V; Johnson TA; Rassenti LZ; Feigal E; Kipps TJ
Favored use of immunoglobulin V(H)4 Genes in AIDS-associated B-cell lymphoma.
Blood 88: 252-60 (1996)

Bevec D; Jaksche H; Oft M; Wohl T; Himmelspach M; Pacher A; Schebesta M; Koettnitz K; Dobrovnik M; Csonga R; Lottspeich F; Hauber J
Inhibition of HIV-1 replication in lymphocytes by mutants of the Rev cofactor eIF-5A.
Science 271: 1858-60 (1996)

Biasolo MA; Radaelli A; Del Pup L; Franchin E; De Giuli-Morghen C; Palu G
A new antisense tRNA construct for the genetic treatment of human immunodeficiency virus type 1 infection.
J Virol 70: 2154-61 (1996)

Bibollet-Ruche F; Galat-Luong A; Cuny G; Sarni-Manchado P; Galat G; Durand JP; Pourrut X; Veas F
Simian immunodeficiency virus infection in a patas monkey (Erythrocebus patas): evidence for cross-species transmission from African green monkeys (Cercopithecus aethiops sabaeus) in the wild.
J Gen Virol 77 (Pt 4): 773-81 (1996)

Bibollet-Ruche F; Galat-Luong A; Cuny G; Sarni-Manchado P; Galat G; Durand JP; Pourrut X; Veas F
Simian immunodeficiency virus infection in a patas monkey (Erythrocebus patas): evidence for cross-species transmission from African green monkeys (Cercopithecus aethiops sabaeus) in the wild.
J Gen Virol 77 (Pt 4): 773-81 (1996)

Bigoni B; Dolcetti R; de Lellis L; Carbone A; Boiocchi M; Cassai E; Di Luca D
Human herpesvirus 8 is present in the lymphoid system of healthy persons and can reactivate in the course of AIDS.
J Infect Dis 173: 542-9 (1996)

Bilello JA; Drusano GL
Relevance of plasma protein binding to antiviral activity and clinical efficacy of inhibitors of human immunodeficiency virus protease [letter]
J Infect Dis 173: 1524-6 (1996)

Bishop JS; Guy-Caffey JK; Ojwang JO; Smith SR; Hogan ME; Cossum PA; Rando RF; Chaudhary N
Intramolecular G-quartet motifs confer nuclease resistance to a potent anti-HIV oligonucleotide.
J Biol Chem 271: 5698-703 (1996)

Biswas DK; Tius MA; Zhuo J; Pardee AB
Canventol inhibits HIV-1 replication by Tat-induced Tar-independent mechanism.
J Acquir Immune Defic Syndr Hum Retrovirol 12: 120-7 (1996)

Blair WS; Fridell RA; Cullen BR
Synergistic enhancement of both initiation and elongation by acidic transcription activation domains.
EMBO J 15: 1658-65 (1996)

Blanco J; Marie I; Callebaut C; Jacotot E; Krust B; Hovanessian AG
Specfic binding of adenosine deaminase but not HIV-1 transactivator protein Tat to human CD26.
Exp Cell Res 225: 102-11 (1996)

Blau J; Xiao H; McCracken S; O'Hare P; Greenblatt J; Bentley D
Three functional classes of transcriptional activation domain.
Mol Cell Biol 16: 2044-55 (1996)

Boerlin P; Boerlin-Petzold F; Goudet J; Durussel C; Pagani JL; Chave JP; Bille J
Typing Candida albicans oral isolates from human immunodeficiency virus-infected patients by multilocus enzyme electrophoresis and DNA fingerprinting.
J Clin Microbiol 34: 1235-48 (1996)

Boerlin P; Boerlin-Petzold F; Goudet J; Durussel C; Pagani JL; Chave JP; Bille J
Typing Candida albicans oral isolates from human immunodeficiency virus-infected patients by multilocus enzyme electrophoresis and DNA fingerprinting.
J Clin Microbiol 34: 1235-48 (1996)

Bolmstedt A; Sjolander S; Hansen JE; Akerblom L; Hemming A; Hu SL; Morein B; Olofsson S
Influence of N-linked glycans in V4-V5 region of human immunodeficiency virus type 1 glycoprotein gp160 on induction of a virus-neutralizing humoral response.
J Acquir Immune Defic Syndr Hum Retrovirol 12: 213-20 (1996)

Boni J
PCR detection of HIV.
Methods Mol Biol 50: 93-107 (1996)

Boni J
PCR detection of HIV.
Methods Mol Biol 50: 93-107 (1996)

Boni J; Pyra H; Schupbach J
Sensitive detection and quantification of particle-associated reverse transcriptase in plasma of HIV-1-infected individuals by the product-enhanced reverse transcriptase (PERT) assay.
J Med Virol 49: 23-8 (1996)

Boni J; Pyra H; Schupbach J
Sensitive detection and quantification of particle-associated reverse transcriptase in plasma of HIV-1-infected individuals by the product-enhanced reverse transcriptase (PERT) assay.
J Med Virol 49: 23-8 (1996)

Borman AM; Paulous S; Clavel F
Resistance of human immunodeficiency virus type 1 to protease inhibitors: selection of resistance mutations in the presence and absence of the drug.
J Gen Virol 77 (Pt 3): 419-26 (1996)

Boudet F; Lecoeur H; Gougeon ML
Apoptosis associated with ex vivo down-regulation of Bcl-2 and up-regulation of Fas in potential cytotoxic CD8+ T lymphocytes during HIV infection.
J Immunol 156: 2282-93 (1996)

Bouhamdan M; Benichou S; Rey F; Navarro JM; Agostini I; Spire B; Camonis J; Slupphaug G; Vigne R; Benarous R; Sire J
Human immunodeficiency virus type 1 Vpr protein binds to the uracil DNA glycosylase DNA repair enzyme.
J Virol 70: 697-704 (1996)

Bour S; Schubert U; Peden K; Strebel K
The envelope glycoprotein of human immunodeficiency virus type 2 enhances viral particle release: a Vpu-like factor?
J Virol 70: 820-9 (1996)

Bourdais J; Biondi R; Sarfati S; Guerreiro C; Lascu I; Janin J; Veron M
Cellular phosphorylation of anti-HIV nucleosides. Role of nucleoside diphosphate kinase.
J Biol Chem 271: 7887-90 (1996)

Bouziane M; Cherny DI; Mouscadet JF; Auclair C
Alternate strand DNA triple helix-mediated inhibition of HIV-1 U5 long terminal repeat integration in vitro.
J Biol Chem 271: 10359-64 (1996)

Boyer PL; Hughes SH
Nucleoside-analogue resistance involves the p66 subunit of HIV-1 RT.
Nat Struct Biol 3: 579-80 (1996)

Braaten D; Franke EK; Luban J
Cyclophilin A is required for an early step in the life cycle of human immunodeficiency virus type 1 before the initiation of reverse transcription.
J Virol 70: 3551-60 (1996)

Braaten D; Franke EK; Luban J
Cyclophilin A is required for the replication of group M human immunodeficiency virus type 1 (HIV-1) and simian immunodeficiency virus SIV(CPZ)GAB but not group O HIV-1 or other primate immunodeficiency viruses.
J Virol 70: 4220-7 (1996)

Braaten D; Franke EK; Luban J
Cyclophilin A is required for the replication of group M human immunodeficiency virus type 1 (HIV-1) and simian immunodeficiency virus SIV(CPZ)GAB but not group O HIV-1 or other primate immunodeficiency viruses.
J Virol 70: 4220-7 (1996)

Brady RL; Barclay AN
The structure of CD4.
Curr Top Microbiol Immunol 205: 1-18 (1996)

Branda RF; Moore AL; Hong R; McCormack JJ; Zon G; Cunningham-Rundles C
B-cell proliferation and differentiation in common variable immunodeficiency patients produced by an antisense oligomer to the rev gene of HIV-1.
Clin Immunol Immunopathol 79: 115-21 (1996)

Branda RF; Moore AL; Lafayette AR; Mathews L; Hong R; Zon G; Brown T; McCormack JJ
Amplification of antibody production by phosphorothioate oligodeoxynucleotides.
J Lab Clin Med 128: 329-38 (1996)

Branda RF; Moore AL; Lafayette AR; Mathews L; Hong R; Zon G; Brown T; McCormack JJ
Amplification of antibody production by phosphorothioate oligodeoxynucleotides.
J Lab Clin Med 128: 329-38 (1996)

Brander C; Corradin G; Hasler T; Pichler WJ
Peptide immunization in humans: a combined CD8+/CD4+ T cell-targeted vaccine restimulates the memory CD4 T cell response but fails to induce cytotoxic T lymphocytes (CTL).
Clin Exp Immunol 105: 18-25 (1996)

Bremer JW; Lew JF; Cooper E; Hillyer GV; Pitt J; Handelsman E; Brambilla D; Moye J; Hoff R
Diagnosis of infection with human immunodeficiency virus type 1 by a DNA polymerase chain reaction assay among infants enrolled in the Women and Infants' Transmission Study [see comments]
J Pediatr 129: 198-207 (1996)

Brenneman DE; Gozes I
A femtomolar-acting neuroprotective peptide [see comments]
J Clin Invest 97: 2299-307 (1996)

Briant L; Coudronniere N; Robert-Hebmann V; Benkirane M; Devaux C
Binding of HIV-1 virions or gp120-anti-gp120 immune complexes to HIV-1-infected quiescent peripheral blood mononuclear cells reveals latent infection.
J Immunol 156: 3994-4004 (1996)

Bridger GJ; Skerlj RT; Padmanabhan S; Martellucci SA; Henson GW; Abrams MJ; Joao HC; Witvrouw M; De Vreese K; Pauwels R; De Clercq E
Synthesis and structure-activity relationships of phenylenebis(methylene)-linked bis-tetraazamacrocycles that inhibit human immunodeficiency virus replication. 2. Effect of heteroaromatic linkers on the activity of bicyclams.
J Med Chem 39: 109-19 (1996)

Bridges EG; Dutschman GE; Gullen EA; Cheng YC
Favorable interaction of beta-L(-) nucleoside analogues with clinically approved anti-HIV nucleoside analogues for the treatment of human immunodeficiency virus.
Biochem Pharmacol 51: 731-6 (1996)

Brogi A; Presentini R; Solazzo D; Piomboni P; Costantino-Ceccarini E
Interaction of human immunodeficiency virus type 1 envelope glycoprotein gp120 with a galactoglycerolipid associated with human sperm.
AIDS Res Hum Retroviruses 12: 483-9 (1996)

Brousset P; Schlaifer D; Roda D; Massip P; Marchou B; Delsol G
Characterization of Epstein-Barr virus-infected cells in benign lymphadenopathy of patients seropositive for human immunodeficiency virus.
Hum Pathol 27: 263-8 (1996)

Brown SA; Imbalzano AN; Kingston RE
Activator-dependent regulation of transcriptional pausing on nucleosomal templates.
Genes Dev 10: 1479-90 (1996)

Broyde S
Setting the stage for predicting RNA thermodynamic properties and their structural components.
Biophys J 70: 1571-2 (1996)

Brunham RC; Kimani J; Bwayo J; Maitha G; Maclean I; Yang C; Shen C; Roman S; Nagelkerke NJ; Cheang M; Plummer FA
The epidemiology of Chlamydia trachomatis within a sexually transmitted diseases core group.
J Infect Dis 173: 950-6 (1996)

Bujia J; Wilmes E; Kastenbauer E; Gurtler L
Influence of chemical allograft preservation procedures on the human immunodeficiency virus.
Laryngoscope 106: 645-7 (1996)

Bukrinskaya AG; Ghorpade A; Heinzinger NK; Smithgall TE; Lewis RE; Stevenson M
Phosphorylation-dependent human immunodeficiency virus type 1 infection and nuclear targeting of viral DNA.
Proc Natl Acad Sci U S A 93: 367-71 (1996)

Calarota S; Jansson M; Levi M; Broliden K; Libonatti O; Wigzell H; Wahren B
Immunodominant glycoprotein 41 epitope identified by seroreactivity in HIV type 1-infected individuals.
AIDS Res Hum Retroviruses 12: 705-13 (1996)

Calarota S; Jansson M; Levi M; Broliden K; Libonatti O; Wigzell H; Wahren B
Immunodominant glycoprotein 41 epitope identified by seroreactivity in HIV type 1-infected individuals.
AIDS Res Hum Retroviruses 12: 705-13 (1996)

Caliendo AM; Savara A; An D; DeVore K; Kaplan JC; D'Aquila RT
Effects of zidovudine-selected human immunodeficiency virus type 1 reverse transcriptase amino acid substitutions on processive DNA synthesis and viral replication.
J Virol 70: 2146-53 (1996)

Callebaut C; Jacotot E; Guichard G; Krust B; Rey-Cuille M; Cointe D; Benkirane N; Blanco J; Muller S; Briand J; Hovanessian AG
Inhibition of HIV infection by pseudopeptides blocking viral envelope glycoprotein-mediated membrane fusion and cell death.
Virology 218: 181-92 (1996)

Callebaut C; Jacotot E; Guichard G; Krust B; Rey-Cuille M; Cointe D; Benkirane N; Blanco J; Muller S; Briand J; Hovanessian AG
Inhibition of HIV infection by pseudopeptides blocking viral envelope glycoprotein-mediated membrane fusion and cell death.
Virology 218: 181-92 (1996)

Camaur D; Trono D
Characterization of human immunodeficiency virus type 1 Vif particle incorporation.
J Virol 70: 6106-11 (1996)

Cameron PU; Lowe MG; Sotzik F; Coughlan AF; Crowe SM; Shortman K
The interaction of macrophage and non-macrophage tropic isolates of HIV-1 with thymic and tonsillar dendritic cells in vitro.
J Exp Med 183: 1851-6 (1996)

Campbell LH; Borg KT; Arrigo SJ
Differential effects of intronic and exonic locations of the human immunodeficiency virus type 1 Rev-responsive element.
Virology 219: 423-31 (1996)

Campiani G; Nacci V; Fiorini I; De Filippis MP; Garofalo A; Greco G; Novellino E; Altamura S; Di Renzo L
Pyrrolobenzothiazepinones and pyrrolobenzoxazepinones: novel and specific non-nucleoside HIV-1 reverse transcriptase inhibitors with antiviral activity.
J Med Chem 39: 2672-80 (1996)

Cannon PM; Byles ED; Kingsman SM; Kingsman AJ
Conserved sequences in the carboxyl terminus of integrase that are essential for human immunodeficiency virus type 1 replication.
J Virol 70: 651-7 (1996)

Canque B; Rosenzwajg M; Gey A; Tartour E; Fridman WH; Gluckman JC
Macrophage inflammatory protein-1alpha is induced by human immunodeficiency virus infection of monocyte-derived macrophages.
Blood 87: 2011-9 (1996)

Cao J; Park IW; Cooper A; Sodroski J
Molecular determinants of acute single-cell lysis by human immunodeficiency virus type 1.
J Virol 70: 1340-54 (1996)

Caputo A; Grossi MP; Bozzini R; Rossi C; Betti M; Marconi PC; Barbanti-Brodano G; Balboni PG
Inhibition of HIV-1 replication and reactivation from latency by tat transdominant negative mutants in the cysteine rich region.
Gene Ther 3: 235-45 (1996)

Cara A; Cereseto A; Lori F; Reitz MS Jr
HIV-1 protein expression from synthetic circles of DNA mimicking the extrachromosomal forms of viral DNA.
J Biol Chem 271: 5393-7 (1996)

Carbone A; Dolcetti R; Gloghini A; Maestro R; Vaccher E; di Luca D; Tirelli U; Boiocchi M
Immunophenotypic and molecular analyses of acquired immune deficiency syndrome-related and Epstein-Barr virus-associated lymphomas: a comparative study.
Hum Pathol 27: 133-46 (1996)

Carr JK; Salminen MO; Koch C; Gotte D; Artenstein AW; Hegerich PA; St Louis D; Burke DS; McCutchan FE
Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand.
J Virol 70: 5935-43 (1996)

Carrillo A; Ratner L
Cooperative effects of the human immunodeficiency virus type 1 envelope variable loops V1 and V3 in mediating infectivity for T cells.
J Virol 70: 1310-6 (1996)

Carrillo A; Ratner L
Human immunodeficiency virus type 1 tropism for T-lymphoid cell lines: role of the V3 loop and C4 envelope determinants.
J Virol 70: 1301-9 (1996)

Cartledge JD; Midgley J; Gazzard BG
Relative growth measurement of Candida species in a single concentration of fluconazole predicts the clinical response to fluconazole in HIV infected patients with oral candidosis.
J Antimicrob Chemother 37: 275-83 (1996)

Catasti P; Bradbury EM; Gupta G
Structure and polymorphism of HIV-1 third variable loops.
J Biol Chem 271: 8236-42 (1996)

Caumont AB; Jamieson GA; Pichuantes S; Nguyen AT; Litvak S; Dupont C
Expression of functional HIV-1 integrase in the yeast Saccharomyces cerevisiae leads to the emergence of a lethal phenotype: potential use for inhibitor screening.
Curr Genet 29: 503-10 (1996)

Cayota A; Vuillier F; Gonzalez G; Dighiero G
In vitro antioxidant treatment recovers proliferative responses of anergic CD4+ lymphocytes from human immunodeficiency virus-infected individuals.
Blood 87: 4746-53 (1996)

Chakrabarti BK; Maitra RK; Ma XZ; Kestler HW
A candidate live inactivatable attenuated vaccine for AIDS.
Proc Natl Acad Sci U S A 93: 9810-5 (1996)

Chakrabarti BK; Maitra RK; Ma XZ; Kestler HW
A candidate live inactivatable attenuated vaccine for AIDS.
Proc Natl Acad Sci U S A 93: 9810-5 (1996)

Chakrabarti BK; Maitra RK; Ma XZ; Kestler HW
A candidate live inactivatable attenuated vaccine for AIDS.
Proc Natl Acad Sci U S A 93: 9810-5 (1996)

Chakrabarti BK; Maitra RK; Ma XZ; Kestler HW
A candidate live inactivatable attenuated vaccine for AIDS.
Proc Natl Acad Sci U S A 93: 9810-5 (1996)

Chen BK; Gandhi RT; Baltimore D
CD4 down-modulation during infection of human T cells with human immunodeficiency virus type 1 involves independent activities of vpu, env, and nef.
J Virol 70: 6044-53 (1996)

Chen H; McBroom DG; Zhu YQ; Gold L; North TW
Inhibitory RNA ligand to reverse transcriptase from feline immunodeficiency virus.
Biochemistry 35: 6923-30 (1996)

Chen P; Cheng PT; Alam M; Beyer BD; Bisacchi GS; Dejneka T; Evans AJ; Greytok JA; Hermsmeier MA; Humphreys WG; Jacobs GA; Kocy O; Lin PF; Lis KA; Marella MA; Ryono DE; Sheaffer AK; Spergel SH; Sun CQ; Tino JA; Vite G; Colonno RJ; Zahler R; Barrish JC
Aminodiol HIV protease inhibitors. Synthesis and structure-activity relationships of P1/P1' compounds: correlation between lipophilicity and cytotoxicity.
J Med Chem 39: 1991-2007 (1996)

Chen X; Bastow K; Goz B; Kucera LS; Morris-Natschke SL; Ishaq KS
Synthesis and evaluation of novel thymidine analogs as antitumor and antiviral agents.
J Med Chem 39: 3412-7 (1996)

Chen Y; Marion PL
Amino acids essential for RNase H activity of hepadnaviruses are also required for efficient elongation of minus-strand viral DNA.
J Virol 70: 6151-6 (1996)

Chen Z; Telfier P; Gettie A; Reed P; Zhang L; Ho DD; Marx PA
Genetic characterization of new West African simian immunodeficiency virus SIVsm: geographic clustering of household-derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop.
J Virol 70: 3617-27 (1996)

Chen Z; Telfier P; Gettie A; Reed P; Zhang L; Ho DD; Marx PA
Genetic characterization of new West African simian immunodeficiency virus SIVsm: geographic clustering of household-derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop.
J Virol 70: 3617-27 (1996)

Chen ZW; Shen L; Regan JD; Kou Z; Ghim SH; Letvin NL
The T cell receptor gene usage by simian immunodeficiency virus gag-specific cytotoxic T lymphocytes in rhesus monkeys.
J Immunol 156: 1469-75 (1996)

Chen ZW; Shen L; Regan JD; Kou Z; Ghim SH; Letvin NL
The T cell receptor gene usage by simian immunodeficiency virus gag-specific cytotoxic T lymphocytes in rhesus monkeys.
J Immunol 156: 1469-75 (1996)

Cherrington JM; Miner R; Hitchcock MJ; Lalezari JP; Drew WL
Susceptibility of human cytomegalovirus to cidofovir is unchanged after limited in vivo exposure to various regimens of drug.
J Infect Dis 173: 987-92 (1996)

Chiang CS; Powell HC; Gold LH; Samimi A; Campbell IL
Macrophage/microglial-mediated primary demyelination and motor disease induced by the central nervous system production of interleukin-3 in transgenic mice.
J Clin Invest 97: 1512-24 (1996)

Chimirri A; Grasso S; Molica C; Monforte AM; Monforte P; Zappala M; Scopelliti R
Anti-HIV agents. IV. Synthesis and in vitro anti-HIV activity of novel 1-(2,6-difluorophenyl)-1H,3H-thiazolo[3,4-a]benzimidazoles.
J Virol 70: 3536-44 (1996)

Chirmule N; Pahwa S
Envelope glycoproteins of human immunodeficiency virus type 1: profound influences on immune functions.
Microbiol Rev 60: 386-406 (1996)

Chirmule N; Pahwa S
Envelope glycoproteins of human immunodeficiency virus type 1: profound influences on immune functions.
Microbiol Rev 60: 386-406 (1996)

Choi YS; Yoshida T; Mimura T; Kaneko Y; Nakashima H; Yamamoto N; Uryu T
Synthesis of sulfated octadecyl ribo-oligosaccharides with potent anti-AIDS virus activity by ring-opening polymerization of a 1,4-anhydroribose derivative.
Carbohydr Res 282: 113-23 (1996)

Choi YS; Yoshida T; Mimura T; Kaneko Y; Nakashima H; Yamamoto N; Uryu T
Synthesis of sulfated octadecyl ribo-oligosaccharides with potent anti-AIDS virus activity by ring-opening polymerization of a 1,4-anhydroribose derivative.
Carbohydr Res 282: 113-23 (1996)

Chou KC
Prediction of human immunodeficiency virus protease cleavage sites in proteins.
Anal Biochem 233: 1-14 (1996)

Chou KC
Prediction of human immunodeficiency virus protease cleavage sites in proteins.
Anal Biochem 233: 1-14 (1996)

Chou KC; Tomasselli AG; Reardon IM; Heinrikson RL
Predicting human immunodeficiency virus protease cleavage sites in proteins by a discriminant function method.
Proteins 24: 51-72 (1996)

Choudhury S; el-Farrash MA; Kuroda MJ; Harada S
Retention of HIV-1 inside infected MOLT-4 cells in association with adhesion-induced cytoskeleton reorganization.
AIDS 10: 363-8 (1996)

Chougnet C; Wynn TA; Clerici M; Landay AL; Kessler HA; Rusnak J; Melcher GP; Sher A; Shearer GM
Molecular analysis of decreased interleukin-12 production in persons infected with human immunodeficiency virus.
J Infect Dis 174: 46-53 (1996)

Chuck S; Grant RM; Katongole-Mbidde E; Conant M; Ganem D
Frequent presence of a novel herpesvirus genome in lesions of human immunodeficiency virus-negative Kaposi's sarcoma.
J Infect Dis 173: 248-51 (1996)

Churchill MJ; Moore JL; Rosenberg M; Brighty DW
The rev-responsive element negatively regulates human immunodeficiency virus type 1 env mRNA expression in primate cells.
J Virol 70: 5786-90 (1996)

Ciernik IF; Berzofsky JA; Carbone DP
Induction of cytotoxic T lymphocytes and antitumor immunity with DNA vaccines expressing single T cell epitopes.
J Immunol 156: 2369-75 (1996)

Civil A; Bakker A; Rensink I; Doerre S; Aarden LA; Verweij CL
Nuclear appearance of a factor that binds the CD28 response element within the interleukin-2 enhancer correlates with interleukin-2 production.
J Biol Chem 271: 8321-7 (1996)

Cleland A; Watson HG; Robertson P; Ludlam CA; Brown AJ
Evolution of zidovudine resistance-associated genotypes in human immunodeficiency virus type 1-infected patients.
J Acquir Immune Defic Syndr Hum Retrovirol 12: 6-18 (1996)

Clements JE; Zink MC
Molecular biology and pathogenesis of animal lentivirus infections.
Clin Microbiol Rev 9: 100-17 (1996)

Clerici M; Balotta C; Salvaggio A; Riva C; Trabattoni D; Papagno L; Berlusconi A; Rusconi S; Villa ML; Moroni M; Galli M
Human immunodeficiency virus (HIV) phenotype and interleukin-2/ interleukin-10 ratio are associated markers of protection and progression in HIV infection.
Blood 88: 574-9 (1996)

Cleveland RP; Liu YC
CD4 Expression by erythroid precursor cells in human bone marrow.
Blood 87: 2275-82 (1996)

Clever JL; Wong ML; Parslow TG
Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA.
J Virol 70: 5902-8 (1996)

Clish CB; Peyton DH; Barklis E
Spectroscopic study of an HIV-1 capsid protein major homology region peptide analog.
FEBS Lett 378: 43-7 (1996)

Codazzi F; Racchetti G; Grohovaz F; Meldolesi J
Transduction signals induced in rat brain cortex astrocytes by the HIV-1 gp120 glycoprotein.
FEBS Lett 384: 135-7 (1996)

Cohen BA
Prognosis and response to therapy of cytomegalovirus encephalitis and meningomyelitis in AIDS.
Neurology 46: 444-50 (1996)

Cohen J
AIDS research. Receptor mutations help slow disease progression [news; comment]
Science 273: 1797-8 (1996)

Cohen J
Likely HIV cofactor found [news; comment]
Science 272: 809-10 (1996)

Cohen J
New role for HIV: a vehicle for moving genes into cells [news; comment]
Science 272: 195 (1996)

Cohen J
AIDS research. Receptor mutations help slow disease progression [news; comment]
Science 273: 1797-8 (1996)

Cohen OJ; Pantaleo G; Holodniy M; Fox CH; Orenstein JM; Schnittman S; Niu M; Graziosi C; Pavlakis GN; Lalezari J; Bartlett JA; Steigbigel RT; Cohn J; Novak R; McMahon D; Bilello J; Fauci AS
Antiretroviral monotherapy in early stage human immunodeficiency virus disease has no detectable effect on virus load in peripheral blood and lymph nodes.
J Infect Dis 173: 849-56 (1996)

Coia G; Hudson PJ; Lilley GG
Construction of recombinant extended single-chain antibody peptide conjugates for use in the diagnosis of HIV-1 and HIV-2.
J Immunol Methods 192: 13-23 (1996)

Colgan J; Yuan HE; Franke EK; Luban J
Binding of the human immunodeficiency virus type 1 Gag polyprotein to cyclophilin A is mediated by the central region of capsid and requires Gag dimerization.
J Virol 70: 4299-310 (1996)

Collette Y; Chang HL; Cerdan C; Chambost H; Algarte M; Mawas C; Imbert J; Burny A; Olive D
Specific Th1 cytokine down-regulation associated with primary clinically derived human immunodeficiency virus type 1 Nef gene-induced expression.
J Immunol 156: 360-70 (1996)

Collette Y; Dutartre H; Benziane A; Romas-Morales; Benarous R; Harris M; Olive D
Physical and functional interaction of Nef with Lck. HIV-1 Nef-induced T-cell signaling defects.
J Biol Chem 271: 6333-41 (1996)

Collette Y; Mawas C; Olive D
Evidence for intact CD28 signaling in T cell hyporesponsiveness induced by the HIV-1 nef gene.
Eur J Immunol 26: 1788-93 (1996)

Collette Y; Mawas C; Olive D
Evidence for intact CD28 signaling in T cell hyporesponsiveness induced by the HIV-1 nef gene.
Eur J Immunol 26: 1788-93 (1996)

Colotto A; Martin I; Ruysschaert JM; Sen A; Hui SW; Epand RM
Structural study of the interaction between the SIV fusion peptide and model membranes.
Biochemistry 35: 980-9 (1996)

Colson P; Lebrun L; Drancourt M; Boue F; Raoult D; Nordmann P
Multiple recurrent bacillary angiomatosis due to Bartonella quintana in an HIV-infected patient [letter]
Eur J Clin Microbiol Infect Dis 15: 178-80 (1996)

Colson P; Lebrun L; Drancourt M; Boue F; Raoult D; Nordmann P
Multiple recurrent bacillary angiomatosis due to Bartonella quintana in an HIV-infected patient [letter]
Eur J Clin Microbiol Infect Dis 15: 178-80 (1996)

Conant K; Ma M; Nath A; Major EO
Extracellular human immunodeficiency virus type 1 Tat protein is associated with an increase in both NF-kappa B binding and protein kinase C activity in primary human astrocytes.
J Virol 70: 1384-9 (1996)

Condos R; McClune A; Rom WN; Schluger NW
Peripheral-blood-based PCR assay to identify patients with active pulmonary tuberculosis.
Lancet 347: 1082-5 (1996)

Cooper JT; Stroka DM; Brostjan C; Palmetshofer A; Bach FH; Ferran C
A20 blocks endothelial cell activation through a NF-kappaB-dependent mechanism.
J Biol Chem 271: 18068-73 (1996)

Corbeil J; Tremblay M; Richman DD
HIV-induced apoptosis requires the CD4 receptor cytoplasmic tail and is accelerated by interaction of CD4 with p56lck.
J Exp Med 183: 39-48 (1996)

Corbellino M; Poirel L; Bestetti G; Pizzuto M; Aubin JT; Capra M; Bifulco C; Berti E; Agut H; Rizzardini G; Galli M; Parravicini C
Restricted tissue distribution of extralesional Kaposi's sarcoma-associated herpesvirus-like DNA sequences in AIDS patients with Kaposi's sarcoma.
AIDS Res Hum Retroviruses 12: 651-7 (1996)

Corbellino M; Poirel L; Bestetti G; Pizzuto M; Aubin JT; Capra M; Bifulco C; Berti E; Agut H; Rizzardini G; Galli M; Parravicini C
Restricted tissue distribution of extralesional Kaposi's sarcoma-associated herpesvirus-like DNA sequences in AIDS patients with Kaposi's sarcoma.
AIDS Res Hum Retroviruses 12: 651-7 (1996)

Cotropia J; Ugen KE; Kliks S; Broliden K; Broliden PA; Hoxie JA; Srikantan V; Williams WV; Weiner DB
A human monoclonal antibody to HIV-1 gp41 with neutralizing activity against diverse laboratory isolates.
J Acquir Immune Defic Syndr Hum Retrovirol 12: 221-32 (1996)

Cotton GJ; Howie SE; Heslop I; Ross JA; Harrison DJ; Ramage R
Design and synthesis of a highly immunogenic, discontinuous epitope of HIV-1 gp120 which binds to CD4+ve transfected cells.
Mol Immunol 33: 171-8 (1996)

Cottrez F; Capron A; Groux H
Selective CD4+ T cell deletion after specific activation in HIV-infected individuals; protection by anti-CD28 monoclonal antibodies.
Clin Exp Immunol 105: 31-8 (1996)

Couldwell DL; Dore GJ; Harkness JL; Marriott DJ; Cooper DA; Edwards R; Li Y; Kaldor JM
Nosocomial outbreak of tuberculosis in an outpatient HIV treatment room.
AIDS 10: 521-5 (1996)

Couldwell DL; Dore GJ; Harkness JL; Marriott DJ; Cooper DA; Edwards R; Li Y; Kaldor JM
Nosocomial outbreak of tuberculosis in an outpatient HIV treatment room.
AIDS 10: 521-5 (1996)

Cullen BR
Virology. New trick from an old foe [news; comment]
Nature 379: 208-9 (1996)

Cunningham AL; Dwyer DE; Mills J; Montagnier L
Structure and function of HIV.
Med J Aust 164: 161-5 (1996)

Cushman M; Golebiewski WM; Graham L; Turpin JA; Rice WG; Fliakas-Boltz V; Buckheit RW Jr
Synthesis and biological evaluation of certain alkenyldiarylmethanes as anti-HIV-1 agents which act as non-nucleoside reverse transcriptase inhibitors.
J Med Chem 39: 3217-27 (1996)

Dada MA; Boshoff CH; Comley MA; Turley H; Schneider JW; Chetty R; Gatter KC
Thymidine phosphorylase expression in Kaposi sarcoma.
J Clin Pathol 49: 400-2 (1996)

Danel K; Larsen E; Pedersen EB; Vestergaard BF; Nielsen C
Synthesis and potent anti-HIV-1 activity of novel 6-benzyluracil analogues of 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine.
J Med Chem 39: 2427-31 (1996)

Danner SA
Zidovudine: anno 1995.
Adv Exp Med Biol 394: 225-43 (1996)

Danner SA
Zidovudine: anno 1995.
Adv Exp Med Biol 394: 225-43 (1996)

De Bernardis F; Chiani P; Ciccozzi M; Pellegrini G; Ceddia T; D'Offizzi G; Quinti I; Sullivan PA; Cassone A
Elevated aspartic proteinase secretion and experimental pathogenicity of Candida albicans isolates from oral cavities of subjects infected with human immunodeficiency virus.
Infect Immun 64: 466-71 (1996)

De Francesco MA; Caruso A; Dima F; Cantalamessa A; Canaris AD; Folghera S; Fiorentini S; Flamminio G; Licenziati S; Peroni L; Gao J; Garotta G; Turano A
IFN-gamma restores HIV- and non-HIV-specific cell mediated immune response in vitro and its activity is neutralized by antibodies from patients with AIDS.
Scand J Immunol 43: 94-100 (1996)

De Luca P; Majello B; Lania L
Sp3 represses transcription when tethered to promoter DNA or targeted to promoter proximal RNA.
J Biol Chem 271: 8533-6 (1996)

DeMartino GN; Proske RJ; Moomaw CR; Strong AA; Song X; Hisamatsu H; Tanaka K; Slaughter CA
Identification, purification, and characterization of a PA700-dependent activator of the proteasome.
J Biol Chem 271: 3112-8 (1996)

Dean M; Carrington M; Winkler C; Huttley GA; Smith MW; Allikmets R; Goedert JJ; Buchbinder SP; Vittinghoff E; Gomperts E; Donfield S; Vlahov D; Kaslow R; Saah A; Rinaldo C; Detels R; O'Brien SJ
Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study [see comments]
Science 273: 1856-62 (1996)

Dean M; Carrington M; Winkler C; Huttley GA; Smith MW; Allikmets R; Goedert JJ; Buchbinder SP; Vittinghoff E; Gomperts E; Donfield S; Vlahov D; Kaslow R; Saah A; Rinaldo C; Detels R; O'Brien SJ
Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study [see comments]
Science 273: 1856-62 (1996)

Debyser Z; De Clercq E
Chemical crosslinking of the subunits of HIV-1 reverse transcriptase.
Protein Sci 5: 278-86 (1996)

Debyser Z; De Clercq E
Chemical crosslinking of the subunits of HIV-1 reverse transcriptase.
Protein Sci 5: 278-86 (1996)

Deckert PM; Ballmaier M; Lang S; Deicher H; Schedel I
CD4-imitating human antibodies in HIV infection and anti-idiotypic vaccination.
J Immunol 156: 826-33 (1996)

Delahunty MD; Rhee I; Freed EO; Bonifacino JS
Mutational analysis of the fusion peptide of the human immunodeficiency virus type 1: identification of critical glycine residues.
Virology 218: 94-102 (1996)

Delahunty MD; Rhee I; Freed EO; Bonifacino JS
Mutational analysis of the fusion peptide of the human immunodeficiency virus type 1: identification of critical glycine residues.
Virology 218: 94-102 (1996)

Demarchi F; d'Adda di Fagagna F; Falaschi A; Giacca M
Activation of transcription factor NF-kappaB by the Tat protein of human immunodeficiency virus type 1.
J Virol 70: 4427-37 (1996)

Denesyuk AI; Matthews S; Zav'yalov VP; Korpela T
Conservative hydrophobic interdomain contacts of IFN-gamma remain in P17 matrix protein of HIV-1.
APMIS 104: 141-6 (1996)

Deng H; Liu R; Ellmeier W; Choe S; Unutmaz D; Burkhart M; Di Marzio P; Marmon S; Sutton RE; Hill CM; Davis CB; Peiper SC; Schall TJ; Littman DR; Landau NR
Identification of a major co-receptor for primary isolates of HIV-1 [see comments]
Nature 381: 661-6 (1996)

Denner J; Persin C; Vogel T; Haustein D; Norley S; Kurth R
The immunosuppressive peptide of HIV-1 inhibits T and B lymphocyte stimulation.
J Acquir Immune Defic Syndr Hum Retrovirol 12: 442-50 (1996)

Deprez B; Sauzet JP; Boutillon C; Martinon F; Tartar A; Sergheraert C; Guillet JG; Gomard E; Gras-Masse H
Comparative efficiency of simple lipopeptide constructs for in vivo induction of virus-specific CTL.
Vaccine 14: 375-82 (1996)

Deprez B; Sauzet JP; Boutillon C; Martinon F; Tartar A; Sergheraert C; Guillet JG; Gomard E; Gras-Masse H
Comparative efficiency of simple lipopeptide constructs for in vivo induction of virus-specific CTL.
Vaccine 14: 375-82 (1996)

Desai NC; Bhatt JJ; Shah BR; Undavia NK; Trivedi PB; Narayanan V
Synthesis of substituted quinazolone derivatives as potential anti-HIV agents (Part III).
Farmaco 51: 361-6 (1996)

Desai NC; Bhatt JJ; Shah BR; Undavia NK; Trivedi PB; Narayanan V
Synthesis of substituted quinazolone derivatives as potential anti-HIV agents (Part III).
Farmaco 51: 361-6 (1996)

Di Somma MM; Majolini MB; Burastero SE; Telford JL; Baldari CT
Cyclosporin A sensitivity of the HIV-1 long terminal repeat identifies distinct p56lck-dependent pathways activated by CD4 triggering.
Eur J Immunol 26: 2181-8 (1996)

Di Somma MM; Majolini MB; Burastero SE; Telford JL; Baldari CT
Cyclosporin A sensitivity of the HIV-1 long terminal repeat identifies distinct p56lck-dependent pathways activated by CD4 triggering.
Eur J Immunol 26: 2181-8 (1996)

Di Stefano M; Gray F; Leitner T; Chiodi F
Analysis of ENV V3 sequences from HIV-1-infected brain indicates restrained virus expression throughout the disease.
J Med Virol 49: 41-8 (1996)

Di Stefano M; Gray F; Leitner T; Chiodi F
Analysis of ENV V3 sequences from HIV-1-infected brain indicates restrained virus expression throughout the disease.
J Med Virol 49: 41-8 (1996)

Di Stefano M; Wilt S; Gray F; Dubois-Dalcq M; Chiodi F
HIV type 1 V3 sequences and the development of dementia during AIDS.
AIDS Res Hum Retroviruses 12: 471-6 (1996)

Diaz JJ; Dodon MD; Schaerer-Uthurralt N; Simonin D; Kindbeiter K; Gazzolo L; Madjar JJ
Post-transcriptional transactivation of human retroviral envelope glycoprotein expression by herpes simplex virus Us11 protein [see comments]
Nature 379: 273-7 (1996)

Diaz L; DeStefano JJ
Strand transfer is enhanced by mismatched nucleotides at the 3' primer terminus: a possible link between HIV reverse transcriptase fidelity and recombination.
Nucleic Acids Res 24: 3086-92 (1996)

Diaz L; DeStefano JJ
Strand transfer is enhanced by mismatched nucleotides at the 3' primer terminus: a possible link between HIV reverse transcriptase fidelity and recombination.
Nucleic Acids Res 24: 3086-92 (1996)

Dickover RE; Garratty EM; Herman SA; Sim MS; Plaeger S; Boyer PJ; Keller M; Deveikis A; Stiehm ER; Bryson YJ
Identification of levels of maternal HIV-1 RNA associated with risk of perinatal transmission. Effect of maternal zidovudine treatment on viral load [see comments]
JAMA 275: 599-605 (1996)

Dittmar MT; Wagener S; Fultz P; Kurth R; Cichutek K
An in vitro assay for acute pathogenicity of immunodeficiency viruses.
Dev Biol Stand 86: 167-73 (1996)

Dittmer U; Spring M; Petry H; Nisslein T; Rieckmann P; Luke W; Stahl-Hennig C; Hunsmann G; Bodemer W
Cell-mediated immune response of macaques immunized with low doses of simian immunodeficiency virus (SIV).
J Biotechnol 44: 105-10 (1996)

Dittmer U; Spring M; Petry H; Nisslein T; Rieckmann P; Luke W; Stahl-Hennig C; Hunsmann G; Bodemer W
Cell-mediated immune response of macaques immunized with low doses of simian immunodeficiency virus (SIV).
J Biotechnol 44: 105-10 (1996)

Dittmer U; Spring M; Petry H; Nisslein T; Rieckmann P; Luke W; Stahl-Hennig C; Hunsmann G; Bodemer W
Cell-mediated immune response of macaques immunized with low doses of simian immunodeficiency virus (SIV).
J Biotechnol 44: 105-10 (1996)

Dolzhanskiy A; Basch RS; Karpatkin S
Development of human megakaryocytes: I. Hematopoietic progenitors (CD34+ bone marrow cells) are enriched with megakaryocytes expressing CD4.
Blood 87: 1353-60 (1996)

Doornenbal P; Seerp Baarsma G; Quint WG; Kijlstra A; Rothbarth PH; Niesters HG
Diagnostic assays in cytomegalovirus retinitis: detection of herpesvirus by simultaneous application of the polymerase chain reaction and local antibody analysis on ocular fluid [see comments]
Br J Ophthalmol 80: 235-40 (1996)

Doranz BJ; Rucker J; Yi Y; Smyth RJ; Samson M; Peiper SC; Parmentier M; Collman RG; Doms RW
A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors.
Cell 85: 1149-58 (1996)

Dorfman T; Gottlinger HG
The human immunodeficiency virus type 1 capsid p2 domain confers sensitivity to the cyclophilin-binding drug SDZ NIM 811.
J Virol 70: 5751-7 (1996)

Dorfman T; Gottlinger HG
The human immunodeficiency virus type 1 capsid p2 domain confers sensitivity to the cyclophilin-binding drug SDZ NIM 811.
J Virol 70: 5751-7 (1996)

Doyon L; Croteau G; Thibeault D; Poulin F; Pilote L; Lamarre D
Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors.
J Virol 70: 3763-9 (1996)

Dragic T; Litwin V; Allaway GP; Martin SR; Huang Y; Nagashima KA; Cayanan C; Maddon PJ; Koup RA; Moore JP; Paxton WA
HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5 [see comments]
Nature 381: 667-73 (1996)

Drobniewski F; Tayler E; Ignatenko N; Paul J; Connolly M; Nye P; Lyagoshina T; Besse C
Tuberculosis in Siberia: 1. An epidemiological and microbiological assessment.
Tuber Lung Dis 77: 199-206 (1996)

Drobniewski F; Tayler E; Ignatenko N; Paul J; Connolly M; Nye P; Lyagoshina T; Besse C
Tuberculosis in Siberia: 1. An epidemiological and microbiological assessment.
Tuber Lung Dis 77: 199-206 (1996)

Drosopoulos WC; Prasad VR
Increased polymerase fidelity of E89G, a nucleoside analog-resistant variant of human immunodeficiency virus type 1 reverse transcriptase.
J Virol 70: 4834-8 (1996)

Du Z; Ilyinskii PO; Sasseville VG; Newstein M; Lackner AA; Desrosiers RC
Requirements for lymphocyte activation by unusual strains of simian immunodeficiency virus.
J Virol 70: 4157-61 (1996)

Duarte EA; Eberl G; Corradin G
Specific tolerization of active cytolytic T lymphocyte responses in vivo with soluble peptides.
Cell Immunol 169: 16-23 (1996)

Dubois V; Lafon ME; Ragnaud JM; Pellegrin JL; Damasio F; Baudouin C; Michaud V; Fleury HJ
Detection of JC virus DNA in the peripheral blood leukocytes of HIV-infected patients.
AIDS 10: 353-8 (1996)

Duenas M; Malmborg AC; Casalvilla R; Ohlin M; Borrebaeck CA
Selection of phage displayed antibodies based on kinetic constants.
Mol Immunol 33: 279-85 (1996)

Dvorakova H; Masojidkova M; Holy A; Balzarini J; Andrei G; Snoeck R; De Clercq E
Synthesis of 2'-aminomethyl derivatives of N-(2-(phosphonomethoxy)ethyl) nucleotide analogues as potential antiviral agents.
J Med Chem 39: 3263-8 (1996)

Ebenbichler CF; Stoiber H; Schneider R; Patsch JR; Dierich MP
The human immunodeficiency virus type 1 transmembrane gp41 protein is a calcium-binding protein and interacts with the putative second-receptor molecules in a calcium-dependent manner.
J Virol 70: 1723-8 (1996)

Eclache V; Magnac C; Pritsch O; Delecluse HJ; Davi F; Raphael M; Dighiero G
Complete nucleotide sequence of Ig V genes in three cases of Burkitt lymphoma associated with AIDS.
Leuk Lymphoma 20: 281-90 (1996)

Ehret A; Westendorp MO; Herr I; Debatin KM; Heeney JL; Frank R; Krammer PH
Resistance of chimpanzee T cells to human immunodeficiency virus type 1 Tat-enhanced oxidative stress and apoptosis.
J Virol 70: 6502-7 (1996)

Ehrhard B; Misselwitz R; Welfle K; Hausdorf G; Glaser RW; Schneider-Mergener J; Welfle H
Chemical modification of recombinant HIV-1 capsid protein p24 leads to the release of a hidden epitope prior to changes of the overall folding of the protein.
Biochemistry 35: 9097-105 (1996)

Ehrlich LS; Fong S; Scarlata S; Zybarth G; Carter C
Partitioning of HIV-1 Gag and Gag-related proteins to membranes.
Biochemistry 35: 3933-43 (1996)

Eich E; Pertz H; Kaloga M; Schulz J; Fesen MR; Mazumder A; Pommier Y
(-)-Arctigenin as a lead structure for inhibitors of human immunodeficiency virus type-1 integrase.
J Med Chem 39: 86-95 (1996)

Ekstrand DH; Awad RJ; Kallander CF; Gronowitz JS
A sensitive assay for the quantification of reverse transcriptase activity based on the use of carrier-bound template and non-radioactive-product detection, with special reference to human-immunodeficiency-virus isolation.
Biotechnol Appl Biochem 23 (Pt 2): 95-105 (1996)

Emiliani S; Van Lint C; Fischle W; Paras P Jr; Ott M; Brady J; Verdin E
A point mutation in the HIV-1 Tat responsive element is associated with postintegration latency.
Proc Natl Acad Sci U S A 93: 6377-81 (1996)

Emini EA; Schleif WA; Deutsch P; Condra JH
In vivo selection of HIV-1 variants with reduced susceptibility to the protease inhibitor L-735,524 and related compounds.
Adv Exp Med Biol 394: 327-31 (1996)

Emini EA; Schleif WA; Deutsch P; Condra JH
In vivo selection of HIV-1 variants with reduced susceptibility to the protease inhibitor L-735,524 and related compounds.
Adv Exp Med Biol 394: 327-31 (1996)

Erickson JW; Burt SK
Structural mechanisms of HIV drug resistance.
Annu Rev Pharmacol Toxicol 36: 545-71 (1996)

Escude C; Giovannangeli C; Sun JS; Lloyd DH; Chen JK; Gryaznov SM; Garestier T; Helene C
Stable triple helices formed by oligonucleotide N3'->P5' phosphoramidates inhibit transcription elongation.
Proc Natl Acad Sci U S A 93: 4365-9 (1996)

Essayag SM; Baily GG; Denning DW; Burnie JP
Karyotyping of fluconazole-resistant yeasts with phenotype reported as Candida krusei or Candida inconspicua.
Int J Syst Bacteriol 46: 35-40 (1996)

Estable MC; Bell B; Merzouki A; Montaner JS; O'Shaughnessy MV; Sadowski IJ
Human immunodeficiency virus type 1 long terminal repeat variants from 42 patients representing all stages of infection display a wide range of sequence polymorphism and transcription activity.
J Virol 70: 4053-62 (1996)

Estaquier J; Tanaka M; Suda T; Nagata S; Golstein P; Ameisen JC
Fas-mediated apoptosis of CD4+ and CD8+ T cells from human immunodeficiency virus-infected persons: differential in vitro preventive effect of cytokines and protease antagonists.
Blood 87: 4959-66 (1996)

Ettmayer P; Billich A; Hecht P; Rosenwirth B; Gstach H
Paracyclophanes: a novel class of water-soluble inhibitors of HIV proteinase.
J Med Chem 39: 3291-9 (1996)

Evers M; Poujade C; Soler F; Ribeill Y; James C; Lelievre Y; Gueguen JC; Reisdorf D; Morize I; Pauwels R; De Clercq E; Henin Y; Bousseau A; Mayaux JF; Le Pecq JB; Dereu N
Betulinic acid derivatives: a new class of human immunodeficiency virus type 1 specific inhibitors with a new mode of action.
J Med Chem 39: 1056-68 (1996)

Faiman GA; Levy R; Anglister J; Horovitz A
Contribution of arginine residues in the RP135 peptide derived from the V3 loop of gp120 to its interaction with the Fv fragment of the 0.5beta HIV-1 neutralizing antibody.
J Biol Chem 271: 13829-33 (1996)

Fan N; Rank KB; Poppe SM; Tarpley WG; Sharma SK
Characterization of the p68/p58 heterodimer of human immunodeficiency virus type 2 reverse transcriptase.
Biochemistry 35: 1911-7 (1996)

Fan N; Rank KB; Slade DE; Poppe SM; Evans DB; Kopta LA; Olmsted RA; Thomas RC; Tarpley WG; Sharma SK
A drug resistance mutation in the inhibitor binding pocket of human immunodeficiency virus type 1 reverse transcriptase impairs DNA synthesis and RNA degradation.
Biochemistry 35: 9737-45 (1996)

Fang G; Weiser B; Visosky AA; Townsend L; Burger H
Molecular cloning of full-length HIV-1 genomes directly from plasma viral RNA.
J Acquir Immune Defic Syndr Hum Retrovirol 12: 352-7 (1996)

Fantuzzi L; Gessani S; Borghi P; Varano B; Conti L; Puddu P; Belardelli F
Induction of interleukin-12 (IL-12) by recombinant glycoprotein gp120 of human immunodeficiency virus type 1 in human monocytes/macrophages: requirement of gamma interferon for IL-12 secretion.
J Virol 70: 4121-4 (1996)

Farnet CM; Wang B; Lipford JR; Bushman FD
Differential inhibition of HIV-1 preintegration complexes and purified integrase protein by small molecules.
Proc Natl Acad Sci U S A 93: 9742-7 (1996)

Farnet CM; Wang B; Lipford JR; Bushman FD
Differential inhibition of HIV-1 preintegration complexes and purified integrase protein by small molecules.
Proc Natl Acad Sci U S A 93: 9742-7 (1996)

Fassler A; Bold G; Capraro HG; Cozens R; Mestan J; Poncioni B; Rosel J; Tintelnot-Blomley M; Lang M
Aza-peptide analogs as potent human immunodeficiency virus type-1 protease inhibitors with oral bioavailability.
J Med Chem 39: 3203-16 (1996)

Fauci AS
Resistance to HIV-1 infection: it's in the genes.
Nat Med 2: 966-7 (1996)

Fauci AS
Resistance to HIV-1 infection: it's in the genes.
Nat Med 2: 966-7 (1996)

Fayolle C; Sebo P; Ladant D; Ullmann A; Leclerc C
In vivo induction of CTL responses by recombinant adenylate cyclase of Bordetella pertussis carrying viral CD8+ T cell epitopes.
J Immunol 156: 4697-706 (1996)

Federico M; Bona R; D'Aloja P; Baiocchi M; Pugliese K; Nappi F; Chelucci C; Mavilio F; Verani P
Anti-HIV viral interference induced by retroviral vectors expressing a nonproducer HIV-1 variant.
Acta Haematol 95: 199-203 (1996)

Fedoroff OYu; Salazar M; Reid BR
Structural variation among retroviral primer-DNA junctions: solution structure of the HIV-1 (-)-strand Okazaki fragment r(gcca)d(CTGC).d(GCAGTGGC).
Biochemistry 35: 11070-80 (1996)

Fedoroff OYu; Salazar M; Reid BR
Structural variation among retroviral primer-DNA junctions: solution structure of the HIV-1 (-)-strand Okazaki fragment r(gcca)d(CTGC).d(GCAGTGGC).
Biochemistry 35: 11070-80 (1996)

Feng Y; Broder CC; Kennedy PE; Berger EA
HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor [see comments]
Science 272: 872-7 (1996)

Filippova IY; Lysogorskaya EN; Anisimova VV; Suvorov LI; Oksenoit ES; Stepanov VM
Fluorogenic peptide substrates for assay of aspartyl proteinases.
Anal Biochem 234: 113-8 (1996)

Filippova IY; Lysogorskaya EN; Anisimova VV; Suvorov LI; Oksenoit ES; Stepanov VM
Fluorogenic peptide substrates for assay of aspartyl proteinases.
Anal Biochem 234: 113-8 (1996)

Finnegan A; Roebuck KA; Nakai BE; Gu DS; Rabbi MF; Song S; Landay AL
IL-10 cooperates with TNF-alpha to activate HIV-1 from latently and acutely infected cells of monocyte/macrophage lineage.
J Immunol 156: 841-51 (1996)

Fitch WM
The variety of human virus evolution.
Mol Phylogenet Evol 5: 247-58 (1996)

Flory E; Hoffmeyer A; Smola U; Rapp UR; Bruder JT
Raf-1 kinase targets GA-binding protein in transcriptional regulation of the human immunodeficiency virus type 1 promoter.
J Virol 70: 2260-8 (1996)

Folgueira L; McElhinny JA; Bren GD; MacMorran WS; Diaz-Meco MT; Moscat J; Paya CV
Protein kinase C-zeta mediates NF-kappa B activation in human immunodeficiency virus-infected monocytes.
J Virol 70: 223-31 (1996)

Fouchier RA; Meyaard L; Brouwer M; Hovenkamp E; Schuitemaker H
Broader tropism and higher cytopathicity for CD4+ T cells of a syncytium-inducing compared to a non-syncytium-inducing HIV-1 isolate as a mechanism for accelerated CD4+ T cell decline in vivo.
Virology 219: 87-95 (1996)

Fowler DA; Rosenthal GJ; Sommadossi JP
Effect of recombinant human hemoglobin on human bone marrow progenitor cells: protection and reversal of 3'-azido-3'-deoxythymidine-induced toxicity.
Toxicol Lett 85: 55-62 (1996)

Frankel DH
Structure of HIV p24 capsid protein revealed [news]
Lancet 348: 184 (1996)

Franzen C; Kuppers R; Muller A; Salzberger B; Fatkenheuer G; Vetten B; Diehl V; Schrappe M
Genetic evidence for latent Septata intestinalis infection in human immunodeficiency virus-infected patients with intestinal microsporidiosis.
J Infect Dis 173: 1038-40 (1996)

Fraziano M; Montesano C; Lombardi VR; Sammarco I; De Pisa F; Mattei M; Valesini G; Pittoni V; Colizzi V
Epitope specificity of anti-HIV antibodies in human and murine autoimmune diseases.
AIDS Res Hum Retroviruses 12: 491-6 (1996)

Freed EO; Martin MA
Domains of the human immunodeficiency virus type 1 matrix and gp41 cytoplasmic tail required for envelope incorporation into virions.
J Virol 70: 341-51 (1996)

Freedman AR; Zhu H; Levine JD; Kalams S; Scadden DT
Generation of human T lymphocytes from bone marrow CD34+ cells in vitro.
Nat Med 2: 46-51 (1996)

Fridell RA; Bogerd HP; Cullen BR
Nuclear export of late HIV-1 mRNAs occurs via a cellular protein export pathway.
Proc Natl Acad Sci U S A 93: 4421-4 (1996)

Fridell RA; Fischer U; Luhrmann R; Meyer BE; Meinkoth JL; Malim MH; Cullen BR
Amphibian transcription factor IIIA proteins contain a sequence element functionally equivalent to the nuclear export signal of human immunodeficiency virus type 1 Rev.
Proc Natl Acad Sci U S A 93: 2936-40 (1996)

Friedland G; Dunkle LW; Cross AP
Stavudine (d4T, Zerit).
Adv Exp Med Biol 394: 271-7 (1996)

Friedland G; Dunkle LW; Cross AP
Stavudine (d4T, Zerit).
Adv Exp Med Biol 394: 271-7 (1996)

Frost RA; Fuhrer J; Steigbigel R; Mariuz P; Lang CH; Gelato MC
Wasting in the acquired immune deficiency syndrome is associated with multiple defects in the serum insulin-like growth factor system.
Clin Endocrinol (Oxf) 44: 501-14 (1996)

Frost RA; Fuhrer J; Steigbigel R; Mariuz P; Lang CH; Gelato MC
Wasting in the acquired immune deficiency syndrome is associated with multiple defects in the serum insulin-like growth factor system.
Clin Endocrinol (Oxf) 44: 501-14 (1996)

Fu GK; Markovitz DM
Purification of the pets factor. A nuclear protein that binds to the inducible TG-rich element of the human immunodeficiency virus type 2 enhancer.
J Biol Chem 271: 19599-605 (1996)

Fuentes GM; Fay PJ; Bambara RA
Relationship between plus strand DNA synthesis removal of downstream segments of RNA by human immunodeficiency virus, murine leukemia virus and avian myeloblastoma virus reverse transcriptases.
Nucleic Acids Res 24: 1719-26 (1996)

Fuentes GM; Rodriguez-Rodriguez L; Palaniappan C; Fay PJ; Bambara RA
Strand displacement synthesis of the long terminal repeats by HIV reverse transcriptase.
J Biol Chem 271: 1966-71 (1996)

Fujii M; Minamino T; Nomura M; Miyamoto KI; Tanaka J; Seiki M
Selective activation of the proto-oncogene c-jun promoter by the transforming protein v-Rel.
Oncogene 12: 2193-202 (1996)

Fujita M; Otsuka M; Sugiura Y
Metal-chelating inhibitors of a zinc finger protein HIV-EP1. Remarkable potentiation of inhibitory activity by introduction of SH groups.
J Med Chem 39: 503-7 (1996)

Furlini G; Vignoli M; Ramazzotti E; Re MC; Visani G; La Placa
A concurrent human herpesvirus-6 infection renders two human hematopoietic progenitor (TF-1 and KG-1) cell lines susceptible to human immunodeficiency virus type-1.
Blood 87: 4737-45 (1996)

Furuishi K; Misumi S; Shoji S
A novel monoclonal antibody to N-myristoyl glycine moiety found a new N-myristoylated HIV-1 p28gag protein in HIV-1-infected cells.
Biochem Biophys Res Commun 222: 344-51 (1996)

Gabrilovich DI; Patterson S; Timofeev AV; Harvey JJ; Knight SC
Mechanism for dendritic cell dysfunction in retroviral infection of mice.
Clin Immunol Immunopathol 80: 139-46 (1996)

Gabrilovich DI; Patterson S; Timofeev AV; Harvey JJ; Knight SC
Mechanism for dendritic cell dysfunction in retroviral infection of mice.
Clin Immunol Immunopathol 80: 139-46 (1996)

Gaidano G; Cechova K; Chang Y; Moore PS; Knowles DM; Dalla-Favera R
Establishment of AIDS-related lymphoma cell lines from lymphomatous effusions.
Leukemia 10: 1237-40 (1996)

Gallay P; Stitt V; Mundy C; Oettinger M; Trono D
Role of the karyopherin pathway in human immunodeficiency virus type 1 nuclear import.
J Virol 70: 1027-32 (1996)

Gao F; Morrison SG; Robertson DL; Thornton CL; Craig S; Karlsson G; Sodroski J; Morgado M; Galvao-Castro B; von Briesen H; Beddows S; Weber J; Sharp PM; Shaw GM; Hahn BH
Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. The WHO and NIAID Networks for HIV Isolation and Characterization.
J Virol 70: 1651-7 (1996)

Garcia S; Dadaglio G; Cilote V; Chenal H; Bondurand A; Gougeon ML
Evidence for an in vivo superantigenic activity in human immunodeficiency virus-infected individuals.
Blood 88: 2151-61 (1996)

Garcia S; Dadaglio G; Cilote V; Chenal H; Bondurand A; Gougeon ML
Evidence for an in vivo superantigenic activity in human immunodeficiency virus-infected individuals.
Blood 88: 2151-61 (1996)

Garza HH Jr; Prakash O; Carr DJ
Aberrant regulation of cytokines in HIV-1 TAT72-transgenic mice.
J Immunol 156: 3631-7 (1996)

Gatignol A; Duarte M; Daviet L; Chang YN; Jeang KT
Sequential steps in Tat trans-activation of HIV-1 mediated through cellular DNA, RNA, and protein binding factors.
Gene Expr 5: 217-28 (1996)

Gatignol A; Duarte M; Daviet L; Chang YN; Jeang KT
Sequential steps in Tat trans-activation of HIV-1 mediated through cellular DNA, RNA, and protein binding factors.
Gene Expr 5: 217-28 (1996)

Gazzinelli RT; Sher A; Cheever A; Gerstberger S; Martin MA; Dickie P
Infection of human immunodeficiency virus 1 transgenic mice with Toxoplasma gondii stimulates proviral transcription in macrophages in vivo.
J Exp Med 183: 1645-55 (1996)

Gebinoga M; Oehlenschlager F
Comparison of self-sustained sequence-replication reaction systems.
Eur J Biochem 235: 256-61 (1996)

Geigenmuller U; Linial ML
Specific binding of human immunodeficiency virus type 1 (HIV-1) Gag-derived proteins to a 5' HIV-1 genomic RNA sequence.
J Virol 70: 667-71 (1996)

Geissler RG; Rossol R; Mentzel U; Ottmann OG; Klein AS; Gute P; Helm EB; Hoelzer D; Ganser A
Gamma delta-T cell-receptor-positive lymphocytes inhibit human hematopoietic progenitor cell growth in HIV type 1-infected patients.
AIDS Res Hum Retroviruses 12: 577-84 (1996)

Geissler RG; Rossol R; Mentzel U; Ottmann OG; Klein AS; Gute P; Helm EB; Hoelzer D; Ganser A
Gamma delta-T cell-receptor-positive lymphocytes inhibit human hematopoietic progenitor cell growth in HIV type 1-infected patients.
AIDS Res Hum Retroviruses 12: 577-84 (1996)

Gerard B; Peponnet C; Brunie G; Cave H; Denamur E; d'Auriol L; Monplaisir N; Simon F; Elion J; Grandchamp B
Fluorometric detection of HIV-1 genome through use of an internal control, inosine-substituted primers, and microtiter plate format.
Clin Chem 42: 696-703 (1996)

Geretti AM; Dings ME; van Els CA; van Baalen CA; Wijnholds FJ; Borleffs JC; Osterhaus AD
Human immunodeficiency virus type 1 (HIV-1)–and Epstein-Barr virus–specific cytotoxic T lymphocyte precursors exhibit different kinetics in HIV-1–infected persons.
J Infect Dis 174: 34-45 (1996)

Ghosh AK; Kincaid JF; Walters DE; Chen Y; Chaudhuri NC; Thompson WJ; Culberson C; Fitzgerald PM; Lee HY; McKee SP; Munson PM; Duong TT; Darke PL; Zugay JA; Schleif WA; Axel MG; Lin J; Huff JR
Nonpeptidal P2 ligands for HIV protease inhibitors: structure-based design, synthesis, and biological evaluation.
J Med Chem 39: 3278-90 (1996)

Ghosh M; Jacques PS; Rodgers DW; Ottman M; Darlix JL; Le Grice SF
Alterations to the primer grip of p66 HIV-1 reverse transcriptase and their consequences for template-primer utilization.
Biochemistry 35: 8553-62 (1996)

Ghosh S; Toth C; Peterlin BM; Seto E
Synergistic activation of transcription by the mutant and wild-type minimal transcriptional activation domain of VP16.
J Biol Chem 271: 9911-8 (1996)

Giacca M; Borella S; Calderazzo F; Bianchi LC; D'Agaro P; Rampazzo C; Bianchi V; Reichard P
Synergistic antiviral action of ribonucleotide reductase inhibitors and 3'-azido-3'-deoxythymidine on HIV type 1 infection in vitro.
AIDS Res Hum Retroviruses 12: 677-82 (1996)

Giacca M; Borella S; Calderazzo F; Bianchi LC; D'Agaro P; Rampazzo C; Bianchi V; Reichard P
Synergistic antiviral action of ribonucleotide reductase inhibitors and 3'-azido-3'-deoxythymidine on HIV type 1 infection in vitro.
AIDS Res Hum Retroviruses 12: 677-82 (1996)

Giavedoni LD; Yilma T
Construction and characterization of replication-competent simian immunodeficiency virus vectors that express gamma interferon.
J Virol 70: 2247-51 (1996)

Gibbs WW
New chip off the old block. Can DNA microprobes do for genetics what microprocessors did for computing? [news]
Sci Am 275: 42, 44 (1996)

Giovannangeli C; Perrouault L; Escude C; Gryaznov S; Helene C
Efficient inhibition of transcription elongation in vitro by oligonucleotide phosphoramidates targeted to proviral HIV DNA.
J Mol Biol 261: 386-98 (1996)

Giovannangeli C; Perrouault L; Escude C; Gryaznov S; Helene C
Efficient inhibition of transcription elongation in vitro by oligonucleotide phosphoramidates targeted to proviral HIV DNA.
J Mol Biol 261: 386-98 (1996)

Giovannangeli C; Perrouault L; Escude C; Nguyen T; Helene C
Specific inhibition of in vitro transcription elongation by triplex-forming oligonucleotide-intercalator conjugates targeted to HIV proviral DNA.
Biochemistry 35: 10539-48 (1996)

Giovannangeli C; Perrouault L; Escude C; Nguyen T; Helene C
Specific inhibition of in vitro transcription elongation by triplex-forming oligonucleotide-intercalator conjugates targeted to HIV proviral DNA.
Biochemistry 35: 10539-48 (1996)

Gitti RK; Lee BM; Walker J; Summers MF; Yoo S; Sundquist WI
Structure of the amino-terminal core domain of the HIV-1 capsid protein.
Science 273: 231-5 (1996)

Glaser RW; Hausdorf G
Binding kinetics of an antibody against HIV p24 core protein measured with real-time biomolecular interaction analysis suggest a slow conformational change in antigen p24.
J Immunol Methods 189: 1-14 (1996)

Gnabre JN; Ito Y; Ma Y; Huang RC
Isolation of anti-HIV-1 lignans from Larrea tridentata by counter-current chromatography.
J Chromatogr A 719: 353-64 (1996)

Goldstein G
HIV-1 Tat protein as a potential AIDS vaccine.
Nat Med 2: 960-4 (1996)

Goldstein G
HIV-1 Tat protein as a potential AIDS vaccine.
Nat Med 2: 960-4 (1996)

Goletti D; Kinter AL; Hardy EC; Poli G; Fauci AS
Modulation of endogenous IL-1 beta and IL-1 receptor antagonist results in opposing effects on HIV expression in chronically infected monocytic cells.
J Immunol 156: 3501-8 (1996)

Gomez CE; Fernandez JR; Iglesias E; Lopez AE; Lobaina L; Noa E; Diaz H; Herrera A; Rolo F; Duarte CA
Complete DNA sequence of the gene encoding the external glycoprotein (gp120) from a Cuban HIV type 1 isolate.
AIDS Res Hum Retroviruses 12: 553-5 (1996)

Gong JH; Zhang M; Modlin RL; Linsley PS; Iyer D; Lin Y; Barnes PF
Interleukin-10 downregulates Mycobacterium tuberculosis-induced Th1 responses and CTLA-4 expression.
Infect Immun 64: 913-8 (1996)

Gonzalez SA; Burny A; Affranchino JL
Identification of domains in the simian immunodeficiency virus matrix protein essential for assembly and envelope glycoprotein incorporation.
J Virol 70: 6384-9 (1996)

Gonzalez SA; Burny A; Affranchino JL
Identification of domains in the simian immunodeficiency virus matrix protein essential for assembly and envelope glycoprotein incorporation.
J Virol 70: 6384-9 (1996)

Goobar-Larsson L; Larsson PT; Debouck C; Towler EM
HIV-1 RT enhances the activity of a tethered dimer of HIV-1 proteinase.
Biochem Biophys Res Commun 220: 203-7 (1996)

Goossens F; De Meester I; Vanhoof G; Scharpe S
Distribution of prolyl oligopeptidase in human peripheral tissues and body fluids.
Eur J Clin Chem Clin Biochem 34: 17-22 (1996)

Gorse GJ; Keefer MC; Belshe RB; Matthews TJ; Forrest BD; Hsieh RH; Koff WC; Hanson CV; Dolin R; Weinhold KJ; Frey SE; Ketter N; Fast PE
A dose-ranging study of a prototype synthetic HIV-1MN V3 branched peptide vaccine. The National Institute of Allergy and Infectious Diseases AIDS Vaccine Evaluation Group.
J Infect Dis 173: 330-9 (1996)

Goudreau G; Carpenter S; Beaulieu N; Jolicoeur P
Vacuolar myelopathy in transgenic mice expressing human immunodeficiency virus type 1 proteins under the regulation of the myelin basic protein gene promoter.
Nat Med 2: 655-61 (1996)

Goulaouic H; Chow SA
Directed integration of viral DNA mediated by fusion proteins consisting of human immunodeficiency virus type 1 integrase and Escherichia coli LexA protein.
J Virol 70: 37-46 (1996)

Graf von Stosch A; Kinzel V; Reed J
Extension of the polarity-dependent "switch phenomenon" of the gp120 binding domain as a target for antiviral chemotherapy.
Biochemistry 35: 411-7 (1996)

Granowitz EV; Saget BM; Angel JB; Wang MZ; Wang A; Dinarello CA; Skolnik PR
Soluble tumor necrosis factor receptors inhibit phorbol myristate acetate and cytokine-induced HIV-1 expression chronically infected U1 cells.
J Acquir Immune Defic Syndr Hum Retrovirol 11: 430-7 (1996)

Gras G; Legendre C; Krzysiek R; Dormont D; Galanaud P; Richard Y
CD40/CD40L interactions and cytokines regulate HIV replication in B cells in vitro.
Virology 220: 309-19 (1996)

Graveley BR; Fleming ES; Gilmartin GM
RNA structure is a critical determinant of poly(A) site recognition by cleavage and polyadenylation specificity factor.
Mol Cell Biol 16: 4942-51 (1996)

Graveley BR; Gilmartin GM
A common mechanism for the enhancement of mRNA 3' processing by U3 sequences in two distantly related lentiviruses.
J Virol 70: 1612-7 (1996)

Graziani A; Galimi F; Medico E; Cottone E; Gramaglia D; Boccaccio C; Comoglio PM
The HIV-1 nef protein interferes with phosphatidylinositol 3-kinase activation 1.
J Biol Chem 271: 6590-3 (1996)

Graziosi C; Gantt KR; Vaccarezza M; Demarest JF; Daucher M; Saag MS; Shaw GM; Quinn TC; Cohen OJ; Welbon CC; Pantaleo G; Fauci AS
Kinetics of cytokine expression during primary human immunodeficiency virus type 1 infection.
Proc Natl Acad Sci U S A 93: 4386-91 (1996)

Grimes PE; Sevall JS; Vojdani A
Cytomegalovirus DNA identified in skin biopsy specimens of patients with vitiligo.
J Am Acad Dermatol 35: 21-6 (1996)

Grzesiek S; Bax A; Clore GM; Gronenborn AM; Hu JS; Kaufman J; Palmer I; Stahl SJ; Wingfield PT
The solution structure of HIV-1 Nef reveals an unexpected fold and permits delineation of the binding surface for the SH3 domain of Hck tyrosine protein kinase.
Nat Struct Biol 3: 340-5 (1996)

Grzesiek S; Stahl SJ; Wingfield PT; Bax A
The CD4 determinant for downregulation by HIV-1 Nef directly binds to Nef. Mapping of the Nef binding surface by NMR.
Biochemistry 35: 10256-61 (1996)

Grzesiek S; Stahl SJ; Wingfield PT; Bax A
The CD4 determinant for downregulation by HIV-1 Nef directly binds to Nef. Mapping of the Nef binding surface by NMR.
Biochemistry 35: 10256-61 (1996)

Gu Z; Li X; Quan Y; Parniak MA; Wainberg MA
Studies of neutralizing monoclonal antibody to human immunodeficiency virus type 1 reverse transcriptase: antagonistic and synergistic effects in reactions performed in the presence of nucleoside and nonnucleoside inhibitors, respectively.
J Virol 70: 2620-6 (1996)

Guillemain C; George F; Courcoul M; Dhiver c; Brunet C; Spire B; Horschowski N; Conciatori M; Sampol J
Monoblastic leukemia in an HIV-infected patient: absence of viral expression in RNA blasts.
Am J Hematol 52: 47-52 (1996)

Guleria I; Teitelbaum R; McAdam RA; Kalpana G; Jacobs WR Jr; Bloom BR
Auxotrophic vaccines for tuberculosis.
Nat Med 2: 334-7 (1996)

Gulizia RJ; Levy JA; Mosier DE
The envelope gp120 gene of human immunodeficiency virus type 1 determines the rate of CD4-positive T-cell depletion in SCID mice engrafted with human peripheral blood leukocytes.
J Virol 70: 4184-7 (1996)

Guptan RC; Thakur V; Sarin SK; Banerjee K; Khandekar P
Frequency and clinical profile of precore and surface hepatitis B mutants in Asian-Indian patients with chronic liver disease [see comments]
Am J Gastroenterol 91: 1312-7 (1996)

Gurtler LG; Zekeng L; Tsague JM; van Brunn A; Afane Ze E; Eberle J; Kaptue L
HIV-1 subtype O: epidemiology, pathogenesis, diagnosis, and perspectives of the evolution of HIV.
Arch Virol Suppl 11: 195-202 (1996)

Gurtler LG; Zekeng L; Tsague JM; van Brunn A; Afane Ze E; Eberle J; Kaptue L
HIV-1 subtype O: epidemiology, pathogenesis, diagnosis, and perspectives of the evolution of HIV.
Arch Virol Suppl 11: 195-202 (1996)

Gussio R; Pattabiraman N; Zaharevitz DW; Kellogg GE; Topol IA; Rice WG; Schaeffer CA; Erickson JW; Burt SK
All-atom models for the non-nucleoside binding site of HIV-1 reverse transcriptase complexed with inhibitors: a 3D QSAR approach.
J Med Chem 39: 1645-50 (1996)

Haas J; Geiss M; Bohler T
False-negative polymerase chain reaction-based diagnosis of human immunodeficiency virus (HIV) type 1 in children infected with HIV strains of African origin [letter]
J Infect Dis 174: 244-5 (1996)

Haddrick M; Lear AL; Cann AJ; Heaphy S
Evidence that a kissing loop structure facilitates genomic RNA dimerisation in HIV-1.
J Mol Biol 259: 58-68 (1996)

Hadgu A
The discrepancy in discrepant analysis.
Lancet 348: 592-3 (1996)

Hadgu A
The discrepancy in discrepant analysis.
Lancet 348: 592-3 (1996)

Hammerschmid F; Billich A; Wasserbauer E; Rosenwirth B
Interactions of HIV-1 proteins with human T-cell cyclophilin A.
Ann N Y Acad Sci 782: 456-61 (1996)

Handley MA; Steigbigel RT; Morrison SA
A role for urokinase-type plasminogen activator in human immunodeficiency virus type 1 infection of macrophages.
J Virol 70: 4451-6 (1996)

Hannongbua S; Lawtrakul L; Limtrakul J
Structure-activity correlation study of HIV-1 inhibitors: electronic and molecular parameters.
J Comput Aided Mol Des 10: 145-52 (1996)

Hannongbua S; Lawtrakul L; Limtrakul J
Structure-activity correlation study of HIV-1 inhibitors: electronic and molecular parameters.
J Comput Aided Mol Des 10: 145-52 (1996)

Hansen JE; Jansson B; Gram GJ; Clausen H; Nielsen JO; Olofsson S
Sensitivity of HIV-1 to neutralization by antibodies against O-linked carbohydrate epitopes despite deletion of O-glycosylation signals in the V3 loop.
Arch Virol 141: 291-300 (1996)

Harada K; Martin SS; Frankel AD
Selection of RNA-binding peptides in vivo.
Nature 380: 175-9 (1996)

Harhaj E; Blaney J; Millhouse S; Sun SC
Differential effects of I kappa B molecules on Tat-mediated transactivation of HIV-1 LTR.
Virology 216: 284-7 (1996)

Harrer E; Harrer T; Barbosa P; Feinberg M; Johnson RP; Buchbinder S; Walker BD
Recognition of the highly conserved YMDD region in the human immunodeficiency virus type 1 reverse transcriptase by HLA-A2-restricted cytotoxic T lymphocytes from an asymptomatic long-term nonprogressor.
J Infect Dis 173: 476-9 (1996)

Harrer T; Harrer E; Kalams SA; Barbosa P; Trocha A; Johnson RP; Elbeik T; Feinberg MB; Buchbinder SP; Walker BD
Cytotoxic T lymphocytes in asymptomatic long-term nonprogressing HIV-1 infection. Breadth and specificity of the response and relation to in vivo viral quasispecies in a person with prolonged infection and low viral load.
J Immunol 156: 2616-23 (1996)

Harrich D; Ulich C; Gaynor RB
A critical role for the TAR element in promoting efficient human immunodeficiency virus type 1 reverse transcription.
J Virol 70: 4017-27 (1996)

Harrigan PR; Kinghorn I; Bloor S; Kemp SD; Najera I; Kohli A; Larder BA
Significance of amino acid variation at human immunodeficiency virus type 1 reverse transcriptase residue 210 for zidovudine susceptibility.
J Virol 70: 5930-4 (1996)

Harrison TS; Levitz SM
Role of IL-12 in peripheral blood mononuclear cell responses to fungi in persons with and without HIV infection.
J Immunol 156: 4492-7 (1996)

Havlir DV; Dube MP; Sattler FR; Forthal DN; Kemper CA; Dunne MW; Parenti DM; Lavelle JP; White ACJr; Witt MD; Bozzette SA; McCutchan JA
Prophylaxis against disseminated Mycobacterium avium complex with weekly azithromycin, daily rifabutin, or both. California Collaborative Treatment Group [see comments]
N Engl J Med 335: 392-8 (1996)

Haynes BF; Pantaleo G; Fauci AS
Toward an understanding of the correlates of protective immunity to HIV infection.
Science 271: 324-8 (1996)

Heimer R; Khoshnood K; Jariwala-Freeman B; Duncan B; Harima Y
Hepatitis in used syringes: the limits of sensitivity of techniques to detect hepatitis B virus (HBV) DNA, hepatitis C virus (HCV) RNA, and antibodies to HBV core and HCV antigens.
J Infect Dis 173: 997-1000 (1996)

Hein J; Stovlbaek J
Combined DNA and protein alignment.
Methods Enzymol 266: 402-18 (1996)

Hein J; Stovlbaek J
Combined DNA and protein alignment.
Methods Enzymol 266: 402-18 (1996)

Heinkelein M; Euler-Konig I; Klinker H; Ruckle-Lanz H; Jassoy C
Lysis of human immunodeficiency virus type 1 antigen-expressing cells by CD4 and CD8 T cells ex vivo.
J Infect Dis 174: 209-13 (1996)

Helbling B; von Overbeck J; Lauterburg BH
Decreased release of glutathione into the systemic circulation of patients with HIV infection.
Eur J Clin Invest 26: 38-44 (1996)

Henderson AJ; Connor RI; Calame KL
C/EBP activators are required for HIV-1 replication and proviral induction in monocytic cell lines.
Immunity 5: 91-101 (1996)

Henderson AJ; Connor RI; Calame KL
C/EBP activators are required for HIV-1 replication and proviral induction in monocytic cell lines.
Immunity 5: 91-101 (1996)

Hendry P; McCall M
Unexpected anisotropy in substrate cleavage rates by asymmetric hammerhead ribozymes.
Nucleic Acids Res 24: 2679-84 (1996)

Herr W; Schneider J; Lohse AW; Meyer zum Buschenfelde KH; Wolfel T
Detection and quantification of blood-derived CD8+ T lymphocytes secreting tumor necrosis factor alpha in response to HLA-A2.1-binding melanoma and viral peptide antigens.
J Immunol Methods 191: 131-42 (1996)

Hertel KJ; Herschlag D; Uhlenbeck OC
Specificity of hammerhead ribozyme cleavage.
EMBO J 15: 3751-7 (1996)

Heusch M; Kraus G; Johnson P; Wong-Staal F
Intracellular immunization against SIVmac utilizing a hairpin ribozyme.
Virology 216: 241-4 (1996)

Heusch M; Kraus G; Johnson P; Wong-Staal F
Intracellular immunization against SIVmac utilizing a hairpin ribozyme.
Virology 216: 241-4 (1996)

Hill AV
HIV and HLA: confusion or complexity? [comment]
Nat Med 2: 395-6 (1996)

Hill CM; Littman DR
Natural resistance to HIV? [news; comment]
Nature 382: 668-9 (1996)

Hill CP; Worthylake D; Bancroft DP; Christensen AM; Sundquist WI
Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly.
Proc Natl Acad Sci U S A 93: 3099-104 (1996)

Himathongkham S; Luciw PA
Restriction of HIV-1 (subtype B) replication at the entry step in rhesus macaque cells.
Virology 219: 485-8 (1996)

Hirsch VM; Fuerst TR; Sutter G; Carroll MW; Yang LC; Goldstein S; Piatak M Jr; Elkins WR; Alvord WG; Montefiori DC; Moss B; Lifson JD
Patterns of viral replication correlate with outcome in simian immunodeficiency virus (SIV)-infected macaques: effect of prior immunization with a trivalent SIV vaccine in modified vaccinia virus Ankara.
J Virol 70: 3741-52 (1996)

Ho WZ; Kaufman D; Song L; Cutillii JR; Douglas SD
Cystamine inhibits human immunodeficiency virus-1 replication in cord blood-derived mononuclear phagocytes and lymphocytes.
Blood 88: 928-33 (1996)

Hodge DR; Robinson L; Watson D; Lautenberger J; Zhang XK; Venanzoni M; Seth A
Interaction of ETS-1 and ERGB/FLI-1 proteins with DNA is modulated by spacing between multiple binding sites as well as phosphorylation.
Oncogene 12: 11-8 (1996)

Hoegl L; Ollert M; Korting HC
The role of Candida albicans secreted aspartic proteinase in the development of candidoses.
J Mol Med 74: 135-42 (1996)

References (1996)

Hoegl L; Ollert M; Korting HC
The role of Candida albicans secreted aspartic proteinase in the development of candidoses.
J Mol Med 74: 135-42 (1996)

Hoffmann JS; Pillaire MJ; Garcia-Estefania D; Lapalu S; Villani G
In vitro bypass replication of the cisplatin-d(GpG) lesion by calf thymus DNA polymerase beta and human immunodeficiency virus type I reverse transcriptase is highly mutagenic.
J Biol Chem 271: 15386-92 (1996)

Hoggard P; Khoo S; Barry M; Back D
Intracellular metabolism of zidovudine and stavudine in combination [letter]
J Infect Dis 174: 671-2 (1996)

Hoggard P; Khoo S; Barry M; Back D
Intracellular metabolism of zidovudine and stavudine in combination [letter]
J Infect Dis 174: 671-2 (1996)

Honda A; Nishimura S
Suppression of translation frameshift by upstream termination codon.
Biochem Biophys Res Commun 221: 602-8 (1996)

Hong L; Treharne A; Hartsuck JA; Foundling S; Tang J
Crystal structures of complexes of a peptidic inhibitor with wild-type and two mutant HIV-1 proteases.
Biochemistry 35: 10627-33 (1996)

Hong L; Treharne A; Hartsuck JA; Foundling S; Tang J
Crystal structures of complexes of a peptidic inhibitor with wild-type and two mutant HIV-1 proteases.
Biochemistry 35: 10627-33 (1996)

Hoog SS; Towler EM; Zhao B; Doyle ML; Debouck C; Abdel-Meguid SS
Human immunodeficiency virus protease ligand specificity conferred by residues outside of the active site cavity.
Biochemistry 35: 10279-86 (1996)

Hoog SS; Towler EM; Zhao B; Doyle ML; Debouck C; Abdel-Meguid SS
Human immunodeficiency virus protease ligand specificity conferred by residues outside of the active site cavity.
Biochemistry 35: 10279-86 (1996)

Hoog SS; Towler EM; Zhao B; Doyle ML; Debouck C; Abdel-Meguid SS
Human immunodeficiency virus protease ligand specificity conferred by residues outside of the active site cavity.
Biochemistry 35: 10279-86 (1996)

Hoog SS; Towler EM; Zhao B; Doyle ML; Debouck C; Abdel-Meguid SS
Human immunodeficiency virus protease ligand specificity conferred by residues outside of the active site cavity.
Biochemistry 35: 10279-86 (1996)

Hoover T; Mikovits J; Court D; Liu YL; Kung HF; Raziuddin
A nuclear matrix-specific factor that binds a specific segment of the negative regulatory element (NRE) of HIV-1 LTR and inhibits NF-kappa(B) activity.
Nucleic Acids Res 24: 1895-900 (1996)

Hopkins AL; Ren J; Esnouf RM; Willcox BE; Jones EY; Ross C; Miyasaka T; Walker RT; Tanaka H; Stammers DK; Stuart DI
Complexes of HIV-1 reverse transcriptase with inhibitors of the HEPT series reveal conformational changes relevant to the design of potent non-nucleoside inhibitors.
J Med Chem 39: 1589-600 (1996)

Hu DJ; Dondero TJ; Rayfield MA; George JR; Schochetman G; Jaffe HW; Luo CC; Kalish ML; Weniger BG; Pau CP; Schable CA; Curran JW
The emerging genetic diversity of HIV. The importance of global surveillance for diagnostics, research, and prevention.
JAMA 275: 210-6 (1996)

Hu R; Oyaizu N; Than S; Kalyanaraman VS; Wang XP; Pahwa S
HIV-1 gp160 induces transforming growth factor-beta production in human PBMC.
Clin Immunol Immunopathol 80: 283-9 (1996)

Hu R; Oyaizu N; Than S; Kalyanaraman VS; Wang XP; Pahwa S
HIV-1 gp160 induces transforming growth factor-beta production in human PBMC.
Clin Immunol Immunopathol 80: 283-9 (1996)

Huang Y; Carmichael GC
Role of polyadenylation in nucleocytoplasmic transport of mRNA.
Mol Cell Biol 16: 1534-42 (1996)

Huang Y; Shalom A; Li Z; Wang J; Mak J; Wainberg MA; Kleiman L
Effects of modifying the tRNA(3Lys) anticodon on the initiation of human immunodeficiency virus type 1 reverse transcription.
J Virol 70: 4700-6 (1996)

Hughes SH; Hostomsky Z; Le Grice SF; Lentz K; Arnold E
What is the orientation of DNA polymerases on their templates?
J Virol 70: 2679-83 (1996)

Humphrey RW; O'Brien TR; Newcomb FM; Nishihara H; Wyvill KM; Ramos GA; Saville MW; Goedert JJ; Straus SE; Yarchoan R
Kaposi's sarcoma (KS)-associated herpesvirus-like DNA sequences in peripheral blood mononuclear cells: association with KS and persistence in patients receiving anti-herpesvirus drugs.
Blood 88: 297-301 (1996)

Huppi K; Siwarski D; Mock BA; Dosik J; Hamel PA
Molecular cloning, chromosomal mapping, and expression of the mouse p107 gene.
Mamm Genome 7: 353-5 (1996)

Huppi K; Siwarski D; Mock BA; Dosik J; Hamel PA
Molecular cloning, chromosomal mapping, and expression of the mouse p107 gene.
Mamm Genome 7: 353-5 (1996)

Hutto C; Zhou Y; He J; Geffin R; Hill M; Scott W; Wood C
Longitudinal studies of viral sequence, viral phenotype, and immunologic parameters of human immunodeficiency virus type 1 infection in perinatally infected twins with discordant disease courses.
J Virol 70: 3589-98 (1996)

Huynen MA; Neumann AU
Rate of killing of HIV-infected T cells and disease progression [letter]
Science 272: 1962 (1996)

Igarashi T; Kuwata T; Takehisa J; Ibuki K; Shibata R; Mukai R; Komatsu T; Adachi A; Ido E; Hayami M
Genomic and biological alteration of a human immunodeficiency virus type 1 (HIV-1)-simian immunodeficiency virus strain mac chimera, with HIV-1 Env, recovered from a long-term carrier monkey.
J Gen Virol 77 (Pt 8): 1649-58 (1996)

Igarashi T; Kuwata T; Takehisa J; Ibuki K; Shibata R; Mukai R; Komatsu T; Adachi A; Ido E; Hayami M
Genomic and biological alteration of a human immunodeficiency virus type 1 (HIV-1)-simian immunodeficiency virus strain mac chimera, with HIV-1 Env, recovered from a long-term carrier monkey.
J Gen Virol 77 (Pt 8): 1649-58 (1996)

Ilves H; Barske C; Junker U; Bohnlein E; Veres G
Retroviral vectors designed for targeted expression of RNA polymerase III-driven transcripts: a comparative study.
Gene 171: 203-8 (1996)

Ilyinskii PO; Desrosiers RC
Efficient transcription and replication of simian immunodeficiency virus in the absence of NF-kappaB and Sp1 binding elements.
J Virol 70: 3118-26 (1996)

Ilyinskii PO; Desrosiers RC
Efficient transcription and replication of simian immunodeficiency virus in the absence of NF-kappaB and Sp1 binding elements.
J Virol 70: 3118-26 (1996)

Imrie A; Carr A; Duncombe C; Finlayson R; Vizzard J; Law M; Kaldor J; Penny R; Cooper DA
Primary infection with zidovudine-resistant human immunodeficiency virus type 1 does not adversely affect outcome at 1 year. Sydney Primary HIV Infection Study Group.
J Infect Dis 174: 195-8 (1996)

Indraccolo S; Mion M; Biagiotti R; Romagnani S; Morfini M; Longo G; Zamarchi R; Chieco-Bianchi L; Amadori A
Genetic variability of the human CD4 V2 domain.
Immunogenetics 44: 70-2 (1996)

Ippolito G; The Studio Italiano Rischio O
Scalpel injury and HIV infection in a surgeon.The Studio Italiano Rischio Occupazionale da HIV (SIROH) [letter]
Lancet 347: 1042 (1996)

Isada CM
Protease inhibitors: promising new weapons against HIV.
Cleve Clin J Med 63: 204-8 (1996)

Isaguliants MG; Kadoshnikov IuP; Kalinina TI; Khuliakov IuE; Semiletov IuA; Smirnov VD; Wahren B
[Expression of HIV-1 epitopes included in particles formed by human hepatitis B virus nucleocapsid protein]
Biokhimiia 61: 532-45 (1996)

Isaguliants MG; Kadoshnikov IuP; Kalinina TI; Khuliakov IuE; Semiletov IuA; Smirnov VD; Wahren B
[Expression of HIV-1 epitopes included in particles formed by human hepatitis B virus nucleocapsid protein]
Biokhimiia 61: 532-45 (1996)

Isel C; Ehresmann C; Ehresmann B; Marquet R
Determining the conformation of RNAs in solution. Application to a retroviral system: structure of the HIV-1 primer binding site region and effect of tRNA(3Lys) binding.
Pharm Acta Helv 71: 11-9 (1996)

Isel C; Lanchy JM; Le Grice SF; Ehresmann C; Ehresmann B; Marquet R
Specific initiation and switch to elongation of human immunodeficiency virus type 1 reverse transcription require the post-transcriptional modifications of primer tRNA3Lys.
EMBO J 15: 917-24 (1996)

Iversen AK; Shafer RW; Wehrly K; Winters MA; Mullins JI; Chesebro B; Merigan TC
Multidrug-resistant human immunodeficiency virus type 1 strains resulting from combination antiretroviral therapy.
J Virol 70: 1086-90 (1996)

Jacobsen H; Haenggi M; Ott M; Duncan IB; Andreoni M; Vella S; Mous J
Reduced sensitivity to saquinavir: an update on genotyping from phase I/II trials.
Antiviral Res 29: 95-7 (1996)

Jacobsen H; Haenggi M; Ott M; Duncan IB; Andreoni M; Vella S; Mous J
Reduced sensitivity to saquinavir: an update on genotyping from phase I/II trials.
Antiviral Res 29: 95-7 (1996)

Jacobsen H; Hanggi M; Ott M; Duncan IB; Owen S; Andreoni M; Vella S; Mous J
In vivo resistance to a human immunodeficiency virus type 1 proteinase inhibitor: mutations, kinetics, and frequencies.
J Infect Dis 173: 1379-87 (1996)

Jacque JM; Fernandez B; Arenzana-Seisdedos F; Thomas D; Baleux F; Virelizier JL; Bachelerie F
Permanent occupancy of the human immunodeficiency virus type 1 enhancer by NF-kappa B is needed for persistent viral replication in monocytes.
J Virol 70: 2930-8 (1996)

Janvier B; Lasarte JJ; Sarobe P; Hoebeke J; Baillou-Beaufils A; Borras-Cuesta F; Barin F
B cell epitopes of HIV type 1 p24 capsid protein: a reassessment.
AIDS Res Hum Retroviruses 12: 519-25 (1996)

Jeffs SA; McKeating J; Lewis S; Craft H; Biram D; Stephens PE; Brady RL
Antigenicity of truncated forms of the human immunodeficiency virus type 1 envelope glycoprotein.
J Gen Virol 77 (Pt 7): 1403-10 (1996)

Jenkins TM; Engelman A; Ghirlando R; Craigie R
A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization.
J Biol Chem 271: 7712-8 (1996)

Jeyaseelan R; Kurabayashi M; Kedes L
Doxorubicin inhibits Tat-dependent transactivation of HIV type 1 LTR.
AIDS Res Hum Retroviruses 12: 569-76 (1996)

Jeyaseelan R; Kurabayashi M; Kedes L
Doxorubicin inhibits Tat-dependent transactivation of HIV type 1 LTR.
AIDS Res Hum Retroviruses 12: 569-76 (1996)

Ji J; Clegg NJ; Peterson KR; Jackson AL; Laird CD; Loeb LA
In vitro expansion of GGC:GCC repeats: identification of the preferred strand of expansion.
Nucleic Acids Res 24: 2835-40 (1996)

Ji X; Klarmann GJ; Preston BD
Effect of human immunodeficiency virus type 1 (HIV-1) nucleocapsid protein on HIV-1 reverse transcriptase activity in vitro.
Biochemistry 35: 132-43 (1996)

Jiang MC; Lin JK; Chen SS
Inhibition of HIV-1 Tat-mediated transactivation by quinacrine and chloroquine.
Biochem Biophys Res Commun 226: 1-7 (1996)

Jiang MC; Lin JK; Chen SS
Inhibition of HIV-1 Tat-mediated transactivation by quinacrine and chloroquine.
Biochem Biophys Res Commun 226: 1-7 (1996)

Joag SV; Li Z; Foresman L; Stephens EB; Zhao LJ; Adany I; Pinson DM; McClure HM; Narayan O
Chimeric simian/human immunodeficiency virus that causes progressive loss of CD4+ T cells and AIDS in pig-tailed macaques.
J Virol 70: 3189-97 (1996)

Joag SV; Li Z; Foresman L; Stephens EB; Zhao LJ; Adany I; Pinson DM; McClure HM; Narayan O
Chimeric simian/human immunodeficiency virus that causes progressive loss of CD4+ T cells and AIDS in pig-tailed macaques.
J Virol 70: 3189-97 (1996)

John S; Marais R; Child R; Light Y; Leonard WJ
Importance of low affinity Elf-1 sites in the regulation of lymphoid-specific inducible gene expression.
J Exp Med 183: 743-50 (1996)

Johnson HM; Torres BA; Soos JM
Superantigens: structure and relevance to human disease.
Proc Soc Exp Biol Med 212: 99-109 (1996)

Johnson VA
Combination therapy for HIV-1 infection-overview: preclinical and clinical analysis of antiretroviral combinations.
Antiviral Res 29: 35-9 (1996)

Johnson VA
Combination therapy for HIV-1 infection-overview: preclinical and clinical analysis of antiretroviral combinations.
Antiviral Res 29: 35-9 (1996)

Jungheim LN; Shepherd TA; Baxter AJ; Burgess J; Hatch SD; Lubbehusen P; Wiskerchen M; Muesing MA
Potent human immunodeficiency virus type 1 protease inhibitors that utilize noncoded D-amino acids as P2/P3 ligands.
J Med Chem 39: 96-108 (1996)

Jurcevic S; Praud C; Coppin HL; Bertrand A; Ricard S; Thomsen M; Lakhdar-Ghazal F; De Preval C
Role of polymorphic residues of human leucocyte antigen-DR molecules on the binding of human immunodeficiency virus peptides.
Immunology 87: 414-20 (1996)

Just JJ; Casabona J; Bertran J; Montane C; Fortuny C; Rodrigo C; Mur A; Bosque M; Jovane L; King MC
MHC class II alleles associated with clinical and immunological manifestations of HIV-1 infection among children in Catalonia, Spain.
Tissue Antigens 47: 313-8 (1996)

Kakizawa J; Ushijima H; Morishita Y; Oka S; Ikeda Y; Muller WE
Diversity of HIV type 1 envelope V3 loop region in saliva.
AIDS Res Hum Retroviruses 12: 561-3 (1996)

Kalams SA; Johnson RP; Dynan MJ; Hartman KE; Harrer T; Harrer E; Trocha AK; Blattner WA; Buchbinder SP; Walker BD
T cell receptor usage and fine specificity of human immunodeficiency virus 1-specific cytotoxic T lymphocyte clones: analysis of quasispecies recognition reveals a dominant response directed against a minor in vivo variant.
J Exp Med 183: 1669-79 (1996)

Kamine J; Elangovan B; Subramanian T; Coleman D; Chinnadurai G
Identification of a cellular protein that specifically interacts with the essential cysteine region of the HIV-1 Tat transactivator.
Virology 216: 357-66 (1996)

Kang MS
Uptake and metabolism of BuCast: a glycoprotein processing inhibitor and a potential anti-HIV drug.
Glycobiology 6: 209-16 (1996)

Kaplan AH; Manchester M; Smith T; Yang YL; Swanstrom R
Conditional human immunodeficiency virus type 1 protease mutants show no role for the viral protease early in virus replication.
J Virol 70: 5840-4 (1996)

Karczewski MK; Strebel K
Cytoskeleton association and virion incorporation of the human immunodeficiency virus type 1 Vif protein.
J Virol 70: 494-507 (1996)

Karlsson GB; Gao F; Robinson J; Hahn B; Sodroski J
Increased envelope spike density and stability are not required for the neutralization resistance of primary human immunodeficiency viruses.
J Virol 70: 6136-42 (1996)

Karsten V; Gordon S; Kirn A; Herbein G
HIV-1 envelope glycoprotein gp120 down-regulates CD4 expression in primary human macrophages through induction of endogenous tumour necrosis factor-alpha.
Immunology 88: 55-60 (1996)

Kashiwada Y; Hashimoto F; Cosentino LM; Chen CH; Garrett PE; Lee KH
Betulinic acid and dihydrobetulinic acid derivatives as potent anti-HIV agents.
J Med Chem 39: 1016-7 (1996)

Kaslow RA; Carrington M; Apple R; Park L; Munoz A; Saah AJ; Goedert JJ; Winkler C; O'Brien SJ; Rinaldo C; Detels R; Blattner W; Phair J; Erlich H; Mann DL
Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection [see comments]
Nat Med 2: 405-11 (1996)

Katlama C; Ingrand D; Loveday C; Clumeck N; Mallolas J; Staszewski S; Johnson M; Hill AM; Pearce G; McDade H
Safety and efficacy of lamivudine-zidovudine combination therapy in antiretroviral-naive patients. A randomized controlled comparison with zidovudine monotherapy. Lamivudine European HIV Working Group.
JAMA 276: 118-25 (1996)

Katzenstein DA; Hammer SM; Hughes MD; Gundacker H; Jackson JB; Fiscus S; Rasheed S; Elbeik T; Reichman R; Japour A; Merigan TC; Hirsch MS
The relation of virologic and immunologic markers to clinical outcomes after nucleoside therapy in HIV-infected adults with 200 to 500 CD4 cells per cubic millimeter. AIDS Clinical Trials Group Study 175 Virology Study Team [see comments]
N Engl J Med 335: 1091-8 (1996)

Katzenstein DA; Hammer SM; Hughes MD; Gundacker H; Jackson JB; Fiscus S; Rasheed S; Elbeik T; Reichman R; Japour A; Merigan TC; Hirsch MS
The relation of virologic and immunologic markers to clinical outcomes after nucleoside therapy in HIV-infected adults with 200 to 500 CD4 cells per cubic millimeter. AIDS Clinical Trials Group Study 175 Virology Study Team [see comments]
N Engl J Med 335: 1091-8 (1996)

Katzman M; Sudol M
Nonspecific alcoholysis, a novel endonuclease activity of human immunodeficiency virus type 1 and other retroviral integrases.
J Virol 70: 2598-604 (1996)

Kaufmann GR; Wenk M; Taeschner W; Peterli B; Gyr K; Meyer UA; Haefeli WE
N-acetyltransferase 2 polymorphism in patients infected with human immunodeficiency virus.
Clin Pharmacol Ther 60: 62-7 (1996)

Kaushal S; La Russa VF; Gartner S; Kessler S; Perfetto S; Yu Z; Ritchey DW; Xu J; Perera P; Kim J; Reid T; Mayers DL; St. Louis D; Mosca JD
Exposure of human CD34+ cells to human immunodeficiency virus type 1 does not influence their expansion and proliferation of hematopoietic progenitors in vitro.
Blood 88: 130-7 (1996)

Kaushik N; Rege N; Yadav PN; Sarafianos SG; Modak MJ; Pandey VN
Biochemical analysis of catalytically crucial aspartate mutants of human immunodeficiency virus type 1 reverse transcriptase.
Biochemistry 35: 11536-46 (1996)

Kaushik N; Rege N; Yadav PN; Sarafianos SG; Modak MJ; Pandey VN
Biochemical analysis of catalytically crucial aspartate mutants of human immunodeficiency virus type 1 reverse transcriptase.
Biochemistry 35: 11536-46 (1996)

Kaye JF; Lever AM
trans-acting proteins involved in RNA encapsidation and viral assembly in human immunodeficiency virus type 1.
J Virol 70: 880-6 (1996)

Kazi S; Cohen PR; Williams F; Schempp R; Reveille JD
The diffuse infiltrative lymphocytosis syndrome. clinical and immunogenetic features in 35 patients.
AIDS 10: 385-91 (1996)

Kazmierski WM; Hazen RJ; Aulabaugh A; StClair MH
Inhibitors of human immunodeficiency virus type 1 derived from gp41 transmembrane protein: structure–activity studies.
J Med Chem 39: 2681-9 (1996)

Keck JL; Marqusee S
The putative substrate recognition loop of Escherichia coli ribonuclease H is not essential for activity.
J Biol Chem 271: 19883-7 (1996)

Keck JL; Marqusee S
The putative substrate recognition loop of Escherichia coli ribonuclease H is not essential for activity.
J Biol Chem 271: 19883-7 (1996)

Keefer MC; Graham BS; McElrath MJ; Matthews TJ; Stablein DM; Corey L; Wright PF; Lawrence D; Fast PE; Weinhold K; Hsieh RH; Chernoff D; Dekker C; Dolin R
Safety and immunogenicity of Env 2-3, a human immunodeficiency virus type 1 candidate vaccine, in combination with a novel adjuvant, MTP-PE/MF59. NIAID AIDS Vaccine Evaluation Group.
AIDS Res Hum Retroviruses 12: 683-93 (1996)

Keefer MC; Graham BS; McElrath MJ; Matthews TJ; Stablein DM; Corey L; Wright PF; Lawrence D; Fast PE; Weinhold K; Hsieh RH; Chernoff D; Dekker C; Dolin R
Safety and immunogenicity of Env 2-3, a human immunodeficiency virus type 1 candidate vaccine, in combination with a novel adjuvant, MTP-PE/MF59. NIAID AIDS Vaccine Evaluation Group.
AIDS Res Hum Retroviruses 12: 683-93 (1996)

Keen NJ; Gait MJ; Karn J
Human immunodeficiency virus type-1 Tat is an integral component of the activated transcription-elongation complex.
Proc Natl Acad Sci U S A 93: 2505-10 (1996)

Kern F; Ode-Hakim S; Vogt K; Hoflich C; Reinke P; Volk HD
The enigma of CD57+CD28- T cell expansion–anergy or activation?
Clin Exp Immunol 104: 180-4 (1996)

Kewalramani VN; Park CS; Gallombardo PA; Emerman M
Protein stability influences human immunodeficiency virus type 2 Vpr virion incorporation and cell cycle effect.
Virology 218: 326-34 (1996)

Kim B; Hathaway TR; Loeb LA
Human immunodeficiency virus reverse transcriptase. Functional mutants obtained by random mutagenesis coupled with genetic selection in Escherichia coli.
J Biol Chem 271: 4872-8 (1996)

Kim CU; McGee LR; Krawczyk SH; Harwood E; Harada Y; Swaminathan S; Bischofberger N; Chen MS; Cherrington JM; Xiong SF; Griffin L; Cundy KC; Lee A; Yu B; Gulnik S; Erickson JW
New series of potent, orally bioavailable, non-peptidic cyclic sulfones as HIV-1 protease inhibitors.
J Med Chem 39: 3431-4 (1996)

Kim CU; McGee LR; Krawczyk SH; Harwood E; Harada Y; Swaminathan S; Bischofberger N; Chen MS; Cherrington JM; Xiong SF; Griffin L; Cundy KC; Lee A; Yu B; Gulnik S; Erickson JW
New series of potent, orally bioavailable, non-peptidic cyclic sulfones as HIV-1 protease inhibitors.
J Med Chem 39: 3431-4 (1996)

Kim FJ; Beeche AA; Hunter JJ; Chin DJ; Hope TJ
Characterization of the nuclear export signal of human T-cell lymphotropic virus type 1 Rex reveals that nuclear export is mediated by position-variable hydrophobic interactions.
Mol Cell Biol 16: 5147-55 (1996)

Kim JH; McLinden RJ; Mosca JD; Burke DS; Boswell RN; Birx DL; Redfield RR
Transcriptional effects of superinfection in HIV chronically infected T cells: studies in dually infected clones.
J Acquir Immune Defic Syndr Hum Retrovirol 12: 329-42 (1996)

Kim JH; McLinden RJ; Mosca JD; Vahey MT; Greene WC; Redfield RR
Inhibition of HIV replication by sense and antisense rev response elements in HIV-based retroviral vectors.
J Acquir Immune Defic Syndr Hum Retrovirol 12: 343-51 (1996)

Kim S; Yu SS; Kim VN
Essential role of NF-kappa B in transactivation of the human immunodeficiency virus long terminal repeat by the human cytomegalovirus 1E1 protein.
J Gen Virol 77 (Pt 1): 83-91 (1996)

Kim SG; Nakashima H; Shoji Y; Inagawa T; Yamamoto N; Kinzuka Y; Takai K; Takaku H
5'-linked lipid-oligodeoxyridonucleotide derivatives as inhibitors of human immunodeficiency virus replication.
Bioorg Med Chem 4: 603-8 (1996)

Kim SG; Nakashima H; Shoji Y; Inagawa T; Yamamoto N; Kinzuka Y; Takai K; Takaku H
5'-linked lipid-oligodeoxyridonucleotide derivatives as inhibitors of human immunodeficiency virus replication.
Bioorg Med Chem 4: 603-8 (1996)

Kimura T; Nishikawa M; Fujisawa J
Uncleaved env gp160 of human immunodeficiency virus type 1 is degraded within the Golgi apparatus but not lysosomes in COS-1 cells.
FEBS Lett 390: 15-20 (1996)

King BL; Vajda S; DeLisi C
Empirical free energy as a target function in docking and design: application to HIV-1 protease inhibitors.
FEBS Lett 384: 87-91 (1996)

King BL; Vajda S; DeLisi C
Empirical free energy as a target function in docking and design: application to HIV-1 protease inhibitors.
FEBS Lett 384: 87-91 (1996)

Kingma DW; Weiss WB; Jaffe ES; Kumar S; Frekko K; Raffeld M
Epstein-Barr virus latent membrane protein-1 oncogene deletions: correlations with malignancy in Epstein-Barr virus–associated lymphoproliferative disorders and malignant lymphomas.
Blood 88: 242-51 (1996)

Kinsey NE; Anderson MG; Unangst TJ; Joag SV; Narayan O; Zink MC; Clements JE
Antigenic variation of SIV: mutations in V4 alter the neutralization profile.
Virology 221: 14-21 (1996)

Kinsey NE; Anderson MG; Unangst TJ; Joag SV; Narayan O; Zink MC; Clements JE
Antigenic variation of SIV: mutations in V4 alter the neutralization profile.
Virology 221: 14-21 (1996)

Kiso Y
Design and synthesis of substrate-based peptidomimetic human immunodeficiency virus protease inhibitors containing the hydroxymethylcarbonyl isostere.
Biopolymers 40: 235-44 (1996)

Kleim JP; Rosner M; Winkler I; Paessens A; Kirsch R; Hsiou Y; Arnold E; Riess G
Selective pressure of a quinoxaline nonnucleoside inhibitor of human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT) on HIV-1 replication results in the emergence of nucleoside RT-inhibitor-specific (RT Leu-74–>Val or Ile and Val-75–>Leu or Ile) HIV-1 mutants.
Proc Natl Acad Sci U S A 93: 34-8 (1996)

Klotman PE; Notkins AL
Transgenic models of human immunodeficiency virus type-1.
Curr Top Microbiol Immunol 206: 197-222 (1996)

Knecht H; Raphael M; McQuain C; Rothenberger S; Pihan G; Camilleri-Broet S; Bachmann E; Kershaw GR; Ryan S; Kittler EL; Quesenberry PJ; Schlaifer D; Woda BA; Brousset P
Deletion variants within the NF-kappa B activation domain of the LMP1 oncogene prevail in acquired immunodeficiency syndrome-related large cell lymphomas and human immunodeficiency virus-negative atypical lymphoproliferations.
Blood 87: 876-81 (1996)

Kollmann TR; Pettoello-Mantovani M; Katopodis NF; Hachamovitch M; Rubinstein A; Kim A; Goldstein H
Inhibition of acute in vivo human immunodeficiency virus infection by human interleukin 10 treatment of SCID mice implanted with human fetal thymus and liver.
Proc Natl Acad Sci U S A 93: 3126-31 (1996)

Kollmus H; Hentze MW; Hauser H
Regulated ribosomal frameshifting by an RNA-protein interaction.
RNA 2: 316-23 (1996)

Kolocouris N; Kolocouris A; Foscolos GB; Fytas G; Neyts J; Padalko E; Balzarini J; Snoeck R; Andrei G; De Clercq E
Synthesis and antiviral activity evaluation of some new aminoadamantane derivatives. 2.
J Med Chem 39: 3307-18 (1996)

Komatsu H; Tsukahara T; Tozawa H
Viral RNA binding properties of human immunodeficiency virus type-2 (HIV-2) nucleocapsid protein-derived synthetic peptides.
Biochem Mol Biol Int 38: 1143-54 (1996)

Komatsu H; Tsukahara T; Tozawa H
Viral RNA binding properties of human immunodeficiency virus type-2 (HIV-2) nucleocapsid protein-derived synthetic peptides.
Biochem Mol Biol Int 38: 1143-54 (1996)

Kondo E; Gottlinger HG
A conserved LXXLF sequence is the major determinant in p6gag required for the incorporation of human immunodeficiency virus type 1 Vpr.
J Virol 70: 159-64 (1996)

Koot M; van 't Wout AB; Kootstra NA; de Goede RE; Tersmette M; Schuitemaker H
Relation between changes in cellular load, evolution of viral phenotype, and the clonal composition of virus populations in the course of human immunodeficiency virus type 1 infection.
J Infect Dis 173: 349-54 (1996)

Kozal MJ; Shah N; Shen N; Yang R; Fucini R; Merigan TC; Richman DD; Morris D; Hubbell E; Chee M; Gingeras TR
Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays.
Nat Med 2: 753-9 (1996)

Kubareva EA; Fedorova OA; Gottikh MB; Tanaka H; Malvy C; Shabarova ZA
NF-kappaB p50 subunit cross-linking to DNA duplexes, containing a monosubstituted pyrophosphate internucleotide bond.
FEBS Lett 381: 35-8 (1996)

Kubota S; Duan L; Furuta RA; Hatanaka M; Pomerantz RJ
Nuclear preservation and cytoplasmic degradation of human immunodeficiency virus type 1 Rev protein.
J Virol 70: 1282-7 (1996)

Kubota S; Hatanaka M; Pomerantz RJ
Nucleo-cytoplasmic redistribution of the HTLV-I Rex protein: alterations by coexpression of the HTLV-I p21x protein.
Virology 220: 502-7 (1996)

Kuiken CL; Cornelissen MT; Zorgdrager F; Hartman S; Gibbs AJ; Goudsmit J
Consistent risk group-associated differences in human immunodeficiency virus type 1 vpr, vpu and V3 sequences despite independent evolution.
J Gen Virol 77 (Pt 4): 783-92 (1996)

Kuipers ME; Huisman JG; Swart PJ; de Bethune MP; Pauwels R; Schuitemaker H; De Clercq E; Meijer DK
Mechanism of anti-HIV activity of negatively charged albumins: biomolecular interaction with the HIV-1 envelope protein gp120.
J Acquir Immune Defic Syndr Hum Retrovirol 11: 419-29 (1996)

Kupiec JJ; Hazebrouck S; Leste-Lasserre T; Sonigo P
Conversion of thymidylate synthase into an HIV protease substrate.
J Biol Chem 271: 18465-70 (1996)

Kurata Si
Sensitization of the HIV-1-LTR upon long term low dose oxidative stress.
J Biol Chem 271: 21798-802 (1996)

Kurata Si
Sensitization of the HIV-1-LTR upon long term low dose oxidative stress.
J Biol Chem 271: 21798-802 (1996)

Kuwabara T; Amontov SV; Warashina M; Ohkawa J; Taira K
Characterization of several kinds of dimer minizyme: simultaneous cleavage at two sites in HIV-1 tat mRNA by dimer minizymes.
Nucleic Acids Res 24: 2302-10 (1996)

Kux A; Bertram S; Hufert FT; Schmitz H; von Laer D
Antibodies to p24 antigen do not specifically detect HIV-infected lymphocytes in AIDS patients.
J Immunol Methods 191: 179-86 (1996)

Kuzmic P; Peranteau AG; Garcia-Echeverria G; Rich DH
Mechanical effects on the kinetics of the HIV proteinase deactivation.
Biochem Biophys Res Commun 221: 313-7 (1996)

Kyostio SR; Owens RA
Identification of mutant adeno-associated virus Rep proteins which are dominant-negative for DNA helicase activity.
Biochem Biophys Res Commun 220: 294-9 (1996)

LaCasse RA; Remington KM; North TW
The mutation frequency of feline immunodeficiency virus enhanced by 3'-azido-3'-deoxythymidine.
J Acquir Immune Defic Syndr Hum Retrovirol 12: 26-32 (1996)

Lachgar A; Bernard J; Bizzini B; Astgen A; Le Coq H; Fouchard M; Chams V; Feldman M; Burny A; Zagury JF
Repair of the in vitro HIV-1-induced immunosuppression and blockade of the generation of functional suppressive CD8 cells by anti-alpha interferon and anti-Tat antibodies.
Biomed Pharmacother 50: 13-8 (1996)

Lafont V; Dornand J; Covassin L; Liautard JP; Favero J
The lectin jacalin triggers CD4-mediated lymphocyte signaling by binding CD4 through a protein-protein interaction.
J Leukoc Biol 59: 691-6 (1996)

Lafrenie RM; Wahl LM; Epstein JS; Hewlett IK; Yamada KM; Dhawan S
HIV-1-Tat modulates the function of monocytes and alters their interactions with microvessel endothelial cells. A mechanism of HIV pathogenesis.
J Immunol 156: 1638-45 (1996)

Laisney IL; Benjamin H; Gefter M; Strosberg AD
Permissive residues within the minimal epitopes of neutralizing monoclonal antibodies to the V3 loop of HIV-1.
Eur J Immunol 26: 1634-40 (1996)

Lal RB; Rudolph DL; Dezzutti CS; Linsley PS; Prince HE
Costimulatory effects of T cell proliferation during infection with human T lymphotropic virus types I and II are mediated through CD80 and CD86 ligands.
J Immunol 157: 1288-96 (1996)

Lam PY; Ru Y; Jadhav PK; Aldrich PE; DeLucca GV; Eyermann CJ; Chang CH; Emmett G; Holler ER; Daneker WF; Li L; Confalone PN; McHugh RJ; Han Q; Li R; Markwalder JA; Seitz SP; Sharpe TR; Bacheler LT; Rayner MM; Klabe RM; Shum L; Winslow DL; Kornhauser DM; Hodge CN; et al
Cyclic HIV protease inhibitors: synthesis, conformational analysis, P2/P2' structure-activity relationship, and molecular recognition of cyclic ureas.
J Med Chem 39: 3514-25 (1996)

Lam PY; Ru Y; Jadhav PK; Aldrich PE; DeLucca GV; Eyermann CJ; Chang CH; Emmett G; Holler ER; Daneker WF; Li L; Confalone PN; McHugh RJ; Han Q; Li R; Markwalder JA; Seitz SP; Sharpe TR; Bacheler LT; Rayner MM; Klabe RM; Shum L; Winslow DL; Kornhauser DM; Hodge CN; et al
Cyclic HIV protease inhibitors: synthesis, conformational analysis, P2/P2' structure-activity relationship, and molecular recognition of cyclic ureas.
J Med Chem 39: 3514-25 (1996)

Lambert DM; Barney S; Lambert AL; Guthrie K; Medinas R; Davis DE; Bucy T; Erickson J; Merutka G; Petteway SR Jr
Peptides from conserved regions of paramyxovirus fusion (F) proteins are potent inhibitors of viral fusion.
Proc Natl Acad Sci U S A 93: 2186-91 (1996)

Lamoril J; Molina JM; de Gouvello A; Garin YJ; Deybach JC; Modai J; Derouin F
Detection by PCR of Toxoplasma gondii in blood in the diagnosis of cerebral toxoplasmosis in patients with AIDS.
J Clin Pathol 49: 89-92 (1996)

Lapham C; Golding B; Inman J; Blackburn R; Manischewitz J; Highet P; Golding H
Brucella abortus conjugated with a peptide derived from the V3 loop of human immunodeficiency virus (HIV) type 1 induces HIV-specific cytotoxic T-cell responses in normal and in CD4+ cell-depleted BALB/c mice.
J Virol 70: 3084-92 (1996)

Larder BA; Kohli A; Bloor S; Kemp SD; Harrigan PR; Schooley RT; Lange JM; Pennington KN; St. Clair MH
Human immunodeficiency virus type 1 drug susceptibility during zidovudine (AZT) monotherapy compared with AZT plus 2',3'-dideoxyinosine or AZT plus 2',3'-dideoxycytidine combination therapy. The protocol 34,225-02 Collaborative Group.
J Virol 70: 5922-9 (1996)

Larsson S; Hotchkiss G; Su J; Kebede T; Andang M; Nyholm T; Johansson B; Sonnerborg A; Vahlne A; Britton S; Ahrlund-Richter L
A novel ribozyme target site located in the HIV-1 nef open reading frame.
Virology 219: 161-9 (1996)

Laskowski RA; Thornton JM; Humblet C; Singh J
X-SITE: use of empirically derived atomic packing preferences to identify favourable interaction regions in the binding sites of proteins.
J Mol Biol 259: 175-201 (1996)

Laughrea M; Jette L
HIV-1 genome dimerization: formation kinetics and thermal stability of dimeric HIV-1Lai RNAs are not improved by the 1-232 and 296-790 regions flanking the kissing-loop domain.
Biochemistry 35: 9366-74 (1996)

Laughrea M; Jette L
Kissing-loop model of HIV-1 genome dimerization: HIV-1 RNAs can assume alternative dimeric forms, and all sequences upstream or downstream of hairpin 248-271 are dispensable for dimer formation.
Biochemistry 35: 1589-98 (1996)

Laurence J; Mitra D; Steiner M; Staiano-Coico L; Jaffe E
Plasma from patients with idiopathic and human immunodeficiency virus-associated thrombotic thrombocytopenic purpura induces apoptosis in microvascular endothelial cells.
Blood 87: 3245-54 (1996)

Lavrik OI; Prasad R; Beard WA; Safronov IV; Dobrikov MI; Srivastava DK; Shishkin GV; Wood TG; Wilson SH
dNTP binding to HIV-1 reverse transcriptase and mammalian DNA polymerase beta as revealed by affinity labeling with a photoreactive dNTP analog.
J Biol Chem 271: 21891-7 (1996)

Lavrik OI; Prasad R; Beard WA; Safronov IV; Dobrikov MI; Srivastava DK; Shishkin GV; Wood TG; Wilson SH
dNTP binding to HIV-1 reverse transcriptase and mammalian DNA polymerase beta as revealed by affinity labeling with a photoreactive dNTP analog.
J Biol Chem 271: 21891-7 (1996)

Law KF; Jagirdar J; Weiden MD; Bodkin M; Rom WN
Tuberculosis in HIV-positive patients: cellular response and immune activation in the lung.
Am J Respir Crit Care Med 153: 1377-84 (1996)

Layton GT; Harris SJ; Myhan J; West D; Gotch F; Hill-Perkins M; Cole JS; Meyers N; Woodrow S; French TJ; Adams SE; Kingsman AJ
Induction of single and dual cytotoxic T-lymphocyte responses to viral proteins in mice using recombinant hybrid Ty-virus-like particles.
Immunology 87: 171-8 (1996)

Leavitt AD; Robles G; Alesandro N; Varmus HE
Human immunodeficiency virus type 1 integrase mutants retain in vitro integrase activity yet fail to integrate viral DNA efficiently during infection.
J Virol 70: 721-8 (1996)

Lech WJ; Wang G; Yang YL; Chee Y; Dorman K; McCrae D; Lazzeroni LC; Erickson JW; Sinsheimer JS; Kaplan AH
In vivo sequence diversity of the protease of human immunodeficiency virus type 1: presence of protease inhibitor-resistant variants in untreated subjects.
J Virol 70: 2038-43 (1996)

Lee CH; Saksela K; Mirza UA; Chait BT; Kuriyan J
Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain.
Cell 85: 931-42 (1996)

Lee MO; Dawson MI; Picard N; Hobbs PD; Pfahl M
A novel class of retinoid antagonists and their mechanism of action.
J Biol Chem 271: 11897-903 (1996)

Lee SP; Han MK
Zinc stimulates Mg2+-dependent 3'-processing activity of human immunodeficiency virus type 1 integrase in vitro.
Biochemistry 35: 3837-44 (1996)

Lee SY; Lee SY; Kandala G; Liou ML; Liou HC; Choi Y
CD30/TNF receptor-associated factor interaction: NF-kappa B activation and binding specificity.
Proc Natl Acad Sci U S A 93: 9699-703 (1996)

Lee SY; Lee SY; Kandala G; Liou ML; Liou HC; Choi Y
CD30/TNF receptor-associated factor interaction: NF-kappa B activation and binding specificity.
Proc Natl Acad Sci U S A 93: 9699-703 (1996)

Lemasson I; Briant L; Hague B; Coudronniere N; Heron L; David C; Rebouissou C; Kindt T; Devaux C
An antibody that binds domain 1 of CD4 inhibits replication of HIV-1, but not HTLV-I, in a CD4-positive/p56lck-negative HTLV-I-transformed cell line.
J Immunol 156: 859-65 (1996)

Levine BL; Mosca JD; Riley JL; Carroll RG; Vahey MT; Jagodzinski LL; Wagner KF; Mayers DL; Burke DS; Weislow OS; St. Louis DC; June CH
Antiviral effect and ex vivo CD4+ T cell proliferation in HIV-positive patients as a result of CD28 costimulation.
Science 272: 1939-43 (1996)

Lewis TL; Matsui SM
Astrovirus ribosomal frameshifting in an infection-transfection transient expression system.
J Virol 70: 2869-75 (1996)

Leydet A; El Hachemi H; Boyer B; Lamaty G; Roque JP; Schols D; Snoeck R; Andrei G; Ikeda S; Neyts J; Reymen D; Este J; Witvrouw M; De Clercq E
Polyanion inhibitors of human immunodeficiency virus and other viruses. Part 2. Polymerized anionic surfactants derived from amino acids and dipeptides.
J Med Chem 39: 1626-34 (1996)

Li MD; Lemke TD; Bronson DL; Faras AJ
Synthesis and analysis of a 640-bp pol region of novel human endogenous retroviral sequences and their evolutionary relationships.
Virology 217: 1-10 (1996)

Li MS; Garcia-Asua G; Bhattacharyya U; Mascagni P; Austen BM; Roberts MM
The Vpr protein of human immunodeficiency virus type 1 binds to nucleocapsid protein p7 in vitro.
Biochem Biophys Res Commun 218: 352-5 (1996)

Li YY; O'Donnell MA; Perez LG
Coexpression of a nonsyncytium inducer HIV-1 glycoprotein inhibits syncytium formation by another HIV-1 Env protein.
Virology 215: 197-202 (1996)

Lima MA; Silveira ER; Marques MS; Santos RH; Gambardela MT
Biologically active flavonoids and terpenoids from Egletes viscosa.
Phytochemistry 41: 217-23 (1996)

Lin L; Daugherty B; Schlom J; Pestka S
Construction of phosphorylatable monoclonal antibody to a tumor-associated antigen.
Cancer Res 56: 4250-4 (1996)

Lin L; Daugherty B; Schlom J; Pestka S
Construction of phosphorylatable monoclonal antibody to a tumor-associated antigen.
Cancer Res 56: 4250-4 (1996)

Lingham RB; Hsu AH; O'Brien JA; Sigmund JM; Sanchez M; Gagliardi MM; Heimbuch BK; Genilloud O; Martin I; Diez MT; Hirsch CF; Zink DL; Liesch JM; Koch GE; Gartner SE; Garrity GM; Tsou NN; Salituro GM
Quinoxapeptins: novel chromodepsipeptide inhibitors of HIV-1 and HIV-2 reverse transcriptase. I. The producing organism and biological activity.
J Antibiot (Tokyo) 49: 253-9 (1996)

Litwin V; Nagashima KA; Ryder AM; Chang CH; Carver JM; Olson WC; Alizon M; Hasel KW; Maddon PJ; Allaway GP
Human immunodeficiency virus type 1 membrane fusion mediated by a laboratory-adapted strain and a primary isolate analyzed by resonance energy transfer.
J Virol 70: 6437-41 (1996)

Liu R; Paxton WA; Choe S; Ceradini D; Martin SR; Horuk R; MacDonald ME; Stuhlmann H; Koup RA; Landau NR
Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection.
Cell 86: 367-77 (1996)

Liu SL; Rodrigo AG; Shankarappa R; Learn GH; Hsu L; Davidov O; Zhao LP; Mullins JI
HIV quasispecies and resampling [letter]
Science 273: 415-6 (1996)

Liu Y; Wang Z; Rana TM
Visualizing a specific contact in the HIV-1 Tat protein fragment and trans-activation responsive region RNA complex by photocross-linking.
J Biol Chem 271: 10391-6 (1996)

Livingstone WJ; Moore M; Innes D; Bell JE; Simmonds P
Frequent infection of peripheral blood CD8-positive T-lymphocytes with HIV-1. Edinburgh Heterosexual Transmission Study Group.
Lancet 348: 649-54 (1996)

Livingstone WJ; Moore M; Innes D; Bell JE; Simmonds P
Frequent infection of peripheral blood CD8-positive T-lymphocytes with HIV-1. Edinburgh Heterosexual Transmission Study Group.
Lancet 348: 649-54 (1996)

Loktev VB; Ilyichev AA; Eroshkin AM; Karpenko LI; Pokrovsky AG; Pereboev AV; Svyatchenko VA; Ignat'ev GM; Smolina MI; Melamed NV; Lebedeva CD; Sandakhchiev LS
Design of immunogens as components of a new generation of molecular vaccines.
J Biotechnol 44: 129-37 (1996)

Loktev VB; Ilyichev AA; Eroshkin AM; Karpenko LI; Pokrovsky AG; Pereboev AV; Svyatchenko VA; Ignat'ev GM; Smolina MI; Melamed NV; Lebedeva CD; Sandakhchiev LS
Design of immunogens as components of a new generation of molecular vaccines.
J Biotechnol 44: 129-37 (1996)

Lorenzo E; Herrera RJ; Lai S; Fischl MA; Hill MD
The Tat and C2-V3 envelope genes in the molecular epidemiology of human immunodeficiency virus-1.
Virology 221: 310-7 (1996)

Lower R; Lower J; Kurth R
The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences.
Proc Natl Acad Sci U S A 93: 5177-84 (1996)

Lu Y; Brosio P; Lafaile M; Li J; Collman RG; Sodroski J; Miller CJ
Vaginal transmission of chimeric simian/human immunodeficiency viruses in rhesus macaques.
J Virol 70: 3045-50 (1996)

Lu Y; Brosio P; Lafaile M; Li J; Collman RG; Sodroski J; Miller CJ
Vaginal transmission of chimeric simian/human immunodeficiency viruses in rhesus macaques.
J Virol 70: 3045-50 (1996)

Lu Y; Salvato MS; Pauza CD; Li J; Sodroski J; Manson K; Wyand M; Letvin N; Jenkins S; Touzjian N; Chutkowski C; Kushner N; LeFaile M; Payne LG; Roberts B
Utility of SHIV for testing HIV-1 vaccine candidates in macaques.
J Acquir Immune Defic Syndr Hum Retrovirol 12: 99-106 (1996)

Lu Y; Salvato MS; Pauza CD; Li J; Sodroski J; Manson K; Wyand M; Letvin N; Jenkins S; Touzjian N; Chutkowski C; Kushner N; LeFaile M; Payne LG; Roberts B
Utility of SHIV for testing HIV-1 vaccine candidates in macaques.
J Acquir Immune Defic Syndr Hum Retrovirol 12: 99-106 (1996)

Luck R; Steger G; Riesner D
Thermodynamic prediction of conserved secondary structure: application to the RRE element of HIV, the tRNA-like element of CMV and the mRNA of prion protein.
J Mol Biol 258: 813-26 (1996)

Ludvigsen A; Werner T; Gimbel W; Erfle V; Brack-Werner R
Down-modulation of HIV-1 LTR activity by an extra-LTR nef gene fragment.
Virology 216: 245-51 (1996)

Luo T; Anderson SJ; Garcia JV
Inhibition of Nef- and phorbol ester-induced CD4 degradation by macrolide antibiotics.
J Virol 70: 1527-34 (1996)

Luo T; Anderson SJ; Garcia JV
Inhibition of Nef- and phorbol ester-induced CD4 degradation by macrolide antibiotics.
J Virol 70: 1527-34 (1996)

Luo T; Garcia JV
The association of Nef with a cellular serine/threonine kinase and its enhancement of infectivity are viral isolate dependent.
J Virol 70: 6493-6 (1996)

Luo T; Garcia JV
The association of Nef with a cellular serine/threonine kinase and its enhancement of infectivity are viral isolate dependent.
J Virol 70: 6493-6 (1996)

Luppi M; Barozzi P; Maiorana A; Artusi T; Trovato R; Marasca R; Savarino M; Ceccherini-Nelli L; Torelli G
Human herpesvirus-8 DNA sequences in human immunodeficiency virus-negative angioimmunoblastic lymphadenopathy and benign lymphadenopathy with giant germinal center hyperplasia and increased vascularity.
Blood 87: 3903-9 (1996)

Lutfey M; Della-Latta P; Kapur V; Palumbo LA; Gurner D; Stotzky G; Brudney K; Dobkin J; Moss A; Musser JM; Kreiswirth BN
Independent origin of mono-rifampin-resistant Mycobacterium tuberculosis in patients with AIDS.
Am J Respir Crit Care Med 153: 837-40 (1996)

Lutz MJ; Held HA; Hottiger M; Hubscher U; Benner SA
Differential discrimination of DNA polymerase for variants of the non-standard nucleobase pair between xanthosine and 2,4-diaminopyrimidine, two components of an expanded genetic alphabet.
Nucleic Acids Res 24: 1308-13 (1996)

Macgrogan G; Vergier B; Dubus P; Beylot-Barry M; Belleannee G; Delaunay MM; Eghbali H; Beylot C; Rivel J; Trojani M; Vital C; De Mascarel A; Bloch B; Merlio JP
CD30-positive cutaneous large cell lymphomas. A comparative study of clinicopathologic and molecular features of 16 cases.
Am J Clin Pathol 105: 440-50 (1996)

Macreadie IG; Arunagiri CK; Hewish DR; White JF; Azad AA
Extracellular addition of a domain of HIV-1 Vpr containing the amino acid sequence motif H(S/F)RIG causes cell membrane permeabilization and death.
Mol Microbiol 19: 1185-92 (1996)

Magnani M; Casabianca A; Fraternale A; Brandi G; Gessani S; Williams R; Giovine M; Damonte G; De Flora A; Benatti U
Synthesis and targeted delivery of an azidothymidine homodinucleotide conferring protection to macrophages against retroviral infection.
Proc Natl Acad Sci U S A 93: 4403-8 (1996)

Maier JA; Mariotti M; Albini A; Comi P; Prat M; Comogilio PM; Soria MR
Over-expression of hepatocyte growth factor in human Kaposi's sarcoma.
Int J Cancer 65: 168-72 (1996)

Mallardo M; Dragonetti E; Baldassarre F; Ambrosino C; Scala G; Quinto I
An NF-kappaB site in the 5'-untranslated leader region of the human immunodeficiency virus type 1 enhances the viral expression in response to NF-kappaB-activating stimuli.
J Biol Chem 271: 20820-7 (1996)

Mallardo M; Dragonetti E; Baldassarre F; Ambrosino C; Scala G; Quinto I
An NF-kappaB site in the 5'-untranslated leader region of the human immunodeficiency virus type 1 enhances the viral expression in response to NF-kappaB-activating stimuli.
J Biol Chem 271: 20820-7 (1996)

Manfredini S; Bazzanini R; Baraldi PG; Bonora M; Marangoni M; Simoni D; Pani A; Scintu F; Pinna E; Pisano L; La Colla P
Pyrazole-related nucleosides. 4. Synthesis and antitumor activity of some 1-tetrahydropyranyl-4-substituted pyrazoles.
Anticancer Drug Des 11: 193-204 (1996)

Manome Y; Yao XJ; Kufe DW; Cohen EA; Fine HA
Selective effects of DNA damaging agents on HIV long terminal repeat activation and virus replication in vitro.
J Acquir Immune Defic Syndr Hum Retrovirol 11: 109-16 (1996)

Margolis-Nunno H; Ben-Hur E; Gottlieb P; Robinson R; Oetjen J; Horowitz B
Inactivation by phthalocyanine photosensitization of multiple forms of human immunodeficiency virus in red cell concentrates.
Transfusion 36: 743-50 (1996)

Margolis-Nunno H; Ben-Hur E; Gottlieb P; Robinson R; Oetjen J; Horowitz B
Inactivation by phthalocyanine photosensitization of multiple forms of human immunodeficiency virus in red cell concentrates.
Transfusion 36: 743-50 (1996)

Markert RL; Ruppach H; Gehring S; Dietrich U; Mierke DF; Kock M; Rubsamen-Waigmann H; Griesinger C
Secondary structural elements as a basis for antibody recognition in the immunodominant region of human immunodeficiency viruses 1 and 2.
Eur J Biochem 237: 188-204 (1996)

Marsh M; Pelchen-Matthews A
Endocytic and exocytic regulation of CD4 expression and function.
Curr Top Microbiol Immunol 205: 107-35 (1996)

Martin I; Schaal H; Scheid A; Ruysschaert JM
Lipid membrane fusion induced by the human immunodeficiency virus type 1 gp41 N-terminal extremity is determined by its orientation in the lipid bilayer.
J Virol 70: 298-304 (1996)

Martin RA; Nayak DP
Mutational analysis of HIV-1 gp160-mediated receptor interference: intracellular complex formation.
Virology 220: 461-72 (1996)

Martin RA; Nayak DP
Membrane anchorage of gp160 is necessary and sufficient to prevent CD4 transport to the cell surface.
Virology 220: 473-9 (1996)

Matsui M; Warburton RJ; Cogswell PC; Baldwin AS Jr; Frelinger JA
Effects of HIV-1 Tat on expression of HLA class I molecules.
J Acquir Immune Defic Syndr Hum Retrovirol 11: 233-40 (1996)

Mavankal G; Ignatius Ou SH; Oliver H; Sigman D; Gaynor RB
Human immunodeficiency virus type 1 and 2 Tat proteins specifically interact with RNA polymerase II.
Proc Natl Acad Sci U S A 93: 2089-94 (1996)

Mazumder A; Neamati N; Sommadossi JP; Gosselin G; Schinazi RF; Imbach JL; Pommier Y
Effects of nucleotide analogues on human immunodeficiency virus type 1 integrase.
Mol Pharmacol 49: 621-8 (1996)

Mazumder A; Wang S; Neamati N; Nicklaus M; Sunder S; Chen J; Milne GW; Rice WG; Burke TR Jr; Pommier Y
Antiretroviral agents as inhibitors of both human immunodeficiency virus type 1 integrase and protease.
J Med Chem 39: 2472-81 (1996)

McBride MS; Panganiban AT
The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures.
J Virol 70: 2963-73 (1996)

McClure HM; Novembre FJ
Simian immunodeficiency virus variants: threat of new lentiviruses.
Am J Med Sci 311: 30-3 (1996)

McClure HM; Novembre FJ
Simian immunodeficiency virus variants: threat of new lentiviruses.
Am J Med Sci 311: 30-3 (1996)

McCutchan FE; Artenstein AW; Sanders-Buell E; Salminen MO; Carr JK; Mascola JR; Yu XF; Nelson KE; Khamboonruang C; Schmitt D; Kieny MP; McNeil JG; Burke DS
Diversity of the envelope glycoprotein among human immunodeficiency virus type 1 isolates of clade E from Asia and Africa.
J Virol 70: 3331-8 (1996)

McElhinny JA; Trushin SA; Bren GD; Chester N; Paya CV
Casein kinase II phosphorylates I kappa B alpha at S-283, S-289, S-293, and T-291 and is required for its degradation.
Mol Cell Biol 16: 899-906 (1996)

McElrath MJ; Corey L; Greenberg PD; Matthews TJ; Montefiori DC; Rowen L; Hood L; Mullins JI
Human immunodeficiency virus type 1 infection despite prior immunization with a recombinant envelope vaccine regimen.
Proc Natl Acad Sci U S A 93: 3972-7 (1996)

McGowan I; Tenant-Flowers M; Jewell DP
Identification of HIV-1 viral RNA in intestinal tissue from patients with early and late HIV infection [letter]
AIDS 10: 548-9 (1996)

McGowan I; Tenant-Flowers M; Jewell DP
Identification of HIV-1 viral RNA in intestinal tissue from patients with early and late HIV infection [letter]
AIDS 10: 548-9 (1996)

McGregor A; Rao MV; Duckworth G; Stockley PG; Connolly BA
Preparation of oligoribonucleotides containing 4-thiouridine using Fpmp chemistry. Photo-crosslinking to RNA binding proteins using 350 nm irradiation.
Nucleic Acids Res 24: 3173-80 (1996)

McGregor A; Rao MV; Duckworth G; Stockley PG; Connolly BA
Preparation of oligoribonucleotides containing 4-thiouridine using Fpmp chemistry. Photo-crosslinking to RNA binding proteins using 350 nm irradiation.
Nucleic Acids Res 24: 3173-80 (1996)

McGuigan C; Cahard D; Sheeka HM; De Clercq E; Balzarini J
Aryl phosphoramidate derivatives of d4T have improved anti-HIV efficacy in tissue culture and may act by the generation of a novel intracellular metabolite.
J Med Chem 39: 1748-53 (1996)

McKeating JA; Zhang YJ; Arnold C; Frederiksson R; Fenyo EM; Balfe P
Chimeric viruses expressing primary envelope glycoproteins of human immunodeficiency virus type I show increased sensitivity to neutralization by human sera.
Virology 220: 450-60 (1996)

McKnight A; Shotton C; Cordell J; Jones I; Simmons G; Clapham PR
Location, exposure, and conservation of neutralizing and nonneutralizing epitopes on human immunodeficiency virus type 2 SU glycoprotein.
J Virol 70: 4598-606 (1996)

McMahon JB; Buckheit RW Jr; Gulakowski RJ; Currens MJ; Vistica DT; Shoemaker RH; Stinson SF; Russell JD; Bader JP; Narayanan VL; Schultz RJ; Brouwer WG; Felauer EE; Boyd MR
Biological and biochemical anti-human immunodeficiency virus activity of UC 38, a new non-nucleoside reverse transcriptase inhibitor.
J Pharmacol Exp Ther 276: 298-305 (1996)

McNeil AJ; Yap PL; Gore SM; Brettle RP; McColl M; Wyld R; Davidson S; Weightman R; Richardson AM; Robertson JR
Association of HLA types A1-B8-DR3 and B27 with rapid and slow progression of HIV disease.
QJM 89: 177-85 (1996)

McNeil AJ; Yap PL; Gore SM; Brettle RP; McColl M; Wyld R; Davidson S; Weightman R; Richardson AM; Robertson JR
Association of HLA types A1-B8-DR3 and B27 with rapid and slow progression of HIV disease.
QJM 89: 177-85 (1996)

Meggio F; D'Agostino DM; Ciminale V; Chieco-Bianchi L; Pinna LA
Phosphorylation of HIV-1 Rev protein: implication of protein kinase CK2 and pro-directed kinases.
Biochem Biophys Res Commun 226: 547-54 (1996)

Meggio F; D'Agostino DM; Ciminale V; Chieco-Bianchi L; Pinna LA
Phosphorylation of HIV-1 Rev protein: implication of protein kinase CK2 and pro-directed kinases.
Biochem Biophys Res Commun 226: 547-54 (1996)

Melekhovets YF; Joshi S
Fusion with an RNA binding domain to confer target RNA specificity to an RNase: design and engineering of Tat-RNase H that specifically recognizes and cleaves HIV-1 RNA in vitro.
Nucleic Acids Res 24: 1908-12 (1996)

Mellors JW
Closing in on human immunodeficiency virus-1.
Nat Med 2: 274-5 (1996)

Melnick M; Reich SH; Lewis KK; Mitchell LJ Jr; Nguyen D; Trippe AJ; Dawson H; Davies JF 2nd; Appelt K; Wu BW; Musick L; Gehlhaar DK; Webber S; Shetty B; Kosa M; Kahil D; Andrada D
Bis tertiary amide inhibitors of the HIV-1 protease generated via protein structure-based iterative design.
J Med Chem 39: 2795-811 (1996)

Mely Y; De Rocquigny H; Morellet N; Roques BP; Gerad D
Zinc binding to the HIV-1 nucleocapsid protein: a thermodynamic investigation by fluorescence spectroscopy.
Biochemistry 35: 5175-82 (1996)

Menu E; Truong TX; Lafon ME; Nguyen TH; Muller-Trutwin MC; Nguyen TT; Deslandres A; Chaouat G; Duong QT; Ha BK; Fleury HJ; Barre-Sinoussi F
HIV type 1 Thai subtype E is predominant in South Vietnam.
AIDS Res Hum Retroviruses 12: 629-33 (1996)

Menu E; Truong TX; Lafon ME; Nguyen TH; Muller-Trutwin MC; Nguyen TT; Deslandres A; Chaouat G; Duong QT; Ha BK; Fleury HJ; Barre-Sinoussi F
HIV type 1 Thai subtype E is predominant in South Vietnam.
AIDS Res Hum Retroviruses 12: 629-33 (1996)

Mesri EA; Cesarman E; Arvanitakis L; Rafii S; Moore MA; Posnett DN; Knowles DM; Asch AS
Human herpesvirus-8/Kaposi's sarcoma-associated herpesvirus is a new transmissible virus that infects B cells.
J Exp Med 183: 2385-90 (1996)

Mestre B; Jakobs A; Pratviel G; Meunier B
Structure/nuclease activity relationships of DNA cleavers based on cationic metalloporphyrin-oligonucleotide conjugates.
Biochemistry 35: 9140-9 (1996)

Metzger AU; Schindler T; Willbold D; Kraft M; Steegborn C; Volkmann A; Frank RW; Rosch P
Structural rearrangements on HIV-1 Tat (32-72) TAR complex formation.
FEBS Lett 384: 255-9 (1996)

Meyer BE; Meinkoth JL; Malim MH
Nuclear transport of human immunodeficiency virus type 1, visna virus, and equine infectious anemia virus Rev proteins: identification of a family of transferable nuclear export signals.
J Virol 70: 2350-9 (1996)

Meyer CF; Wang X; Chang C; Templeton D; Tan TH
Interaction between c-Rel and the mitogen-activated protein kinase kinase kinase 1 signaling cascade in mediating kappaB enhancer activation.
J Biol Chem 271: 8971-6 (1996)

Michienzi A; Prislei S; Bozzoni I
U1 small nuclear RNA chimeric ribozymes with substrate specificity for the Rev pre-mRNA of human immunodeficiency virus.
Proc Natl Acad Sci U S A 93: 7219-24 (1996)

Miedema F; Klein MR
AIDS pathogenesis: a finite immune response to blame? [comment]
Science 272: 505-6 (1996)

Miele G; Lever AM
Expression of mutant and wild-type gag proteins for gene therapy in HIV-1 infection.
Gene Ther 3: 357-61 (1996)

Miele G; Lever AM
Expression of mutant and wild-type gag proteins for gene therapy in HIV-1 infection.
Gene Ther 3: 357-61 (1996)

Miele G; Mouland A; Harrison GP; Cohen E; Lever AM
The human immunodeficiency virus type 1 5' packaging signal structure affects translation but does not function as an internal ribosome entry site structure.
J Virol 70: 944-51 (1996)

Mikaelian I; Krieg M; Gait MJ; Karn J
Interactions of INS (CRS) elements and the splicing machinery regulate the production of Rev-responsive mRNAs.
J Mol Biol 257: 246-64 (1996)

Miller MD; Greene WC
Is the Nef protein of HIV-1 required for pathogenesis?
Trends Microbiol 4: 171-2; discussion 173 (1996)

Miller MD; Greene WC
Is the Nef protein of HIV-1 required for pathogenesis?
Trends Microbiol 4: 171-2; discussion 173 (1996)

Mindell DP
Positive selection and rates of evolution in immunodeficiency viruses from humans and chimpanzees.
Proc Natl Acad Sci U S A 93: 3284-8 (1996)

Mindell DP
Positive selection and rates of evolution in immunodeficiency viruses from humans and chimpanzees.
Proc Natl Acad Sci U S A 93: 3284-8 (1996)

Mitra D; Steiner M; Lynch DH; Staiano-Coico L; Laurence J
HIV-1 upregulates Fas ligand expression in CD4+ T cells in vitro and in vivo: association with Fas-mediated apoptosis and modulation by aurintricarboxylic acid.
Immunology 87: 581-5 (1996)

Miura T; Shibata R; Adachi A; Kuwata T; Chen J; Jin M; Ido E; Hayami M
Genetic complementation between replication-defective mutants of HIV-1 and SIVagm.
Arch Virol 141: 31-41 (1996)

Miura T; Shibata R; Adachi A; Kuwata T; Chen J; Jin M; Ido E; Hayami M
Genetic complementation between replication-defective mutants of HIV-1 and SIVagm.
Arch Virol 141: 31-41 (1996)

Molla A; Korneyeva M; Gao Q; Vasavanonda S; Schipper PJ; Mo HM; Markowitz M; Chernyavskiy T; Niu P; Lyons N; Hsu A; Granneman GR; Ho DD; Boucher CA; Leonard JM; Norbeck DW; Kempf DJ
Ordered accumulation of mutations in HIV protease confers resistance to ritonavir.
Nat Med 2: 760-6 (1996)

Momany C; Kovari LC; Prongay AJ; Keller W; Gitti RK; Lee BM; Gorbalenya AE; Tong L; McClure J; Ehrlich LS; Summers MF; Carter C; Rossmann MG
Crystal structure of dimeric HIV-1 capsid protein.
Nat Struct Biol 3: 763-70 (1996)

Momany C; Kovari LC; Prongay AJ; Keller W; Gitti RK; Lee BM; Gorbalenya AE; Tong L; McClure J; Ehrlich LS; Summers MF; Carter C; Rossmann MG
Crystal structure of dimeric HIV-1 capsid protein.
Nat Struct Biol 3: 763-70 (1996)

Monini P; Rotola A; de Lellis L; Corallini A; Secchiero P; Albini A; Benelli R; Parravicini C; Barbanti-Brodano G; Cassai E
Latent BK virus infection and Kaposi's sarcoma pathogenesis.
Int J Cancer 66: 717-22 (1996)

Moore JP; Cao Y; Leu J; Qin L; Korber B; Ho DD
Inter- and intraclade neutralization of human immunodeficiency virus type 1: genetic clades do not correspond to neutralization serotypes but partially correspond to gp120 antigenic serotypes.
J Virol 70: 427-44 (1996)

Moore JP; Koup RA
Chemoattractants attract HIV researchers [comment]
J Exp Med 184: 311-3 (1996)

Moore PS; Gao SJ; Dominguez G; Cesarman E; Lungu O; Knowles DM; Garber R; Pellett PE; McGeoch DJ; Chang Y
Primary characterization of a herpesvirus agent associated with Kaposi's sarcoma.
J Virol 70: 549-58 (1996)

Mori S; Murakami-Mori K; Jewett A; Nakamura S; Bonavida B
Resistance of AIDS-associated Kaposi's sarcoma cells to Fas-mediated apoptosis.
Cancer Res 56: 1874-9 (1996)

Morimoto M; Otake T; Mori H; Kawahata T; Ueba N; Okubo S; Yasunaga K; Sano K; Nakano T; Nakai M
[Prognosis and evaluation of drug therapy by V3 and RT gene analysis of HIV-1]
Kansenshogaku Zasshi 70: 347-53 (1996)

Morse JH; Barst RJ; Itescu S; Flaster ER; Sinha G; Zhang Y; Fotino M
Primary pulmonary hypertension in HIV infection: an outcome determined by particular HLA class II alleles.
Am J Respir Crit Care Med 153: 1299-301 (1996)

Moses AV; Stenglein SG; Strussenberg JG; Wehrly K; Chesebro B; Nelson JA
Sequences regulating tropism of human immunodeficiency virus type 1 for brain capillary endothelial cells map to a unique region on the viral genome.
J Virol 70: 3401-6 (1996)

Mossman SP; Bex F; Berglund P; Arthos J; O'Neil SP; Riley D; Maul DH; Bruck C; Momin P; Burny A; Fultz PN; Mullins JI; Liljestrom P; Hoover EA
Protection against lethal simian immunodeficiency virus SIVsmmPBj14 disease by a recombinant Semliki Forest virus gp160 vaccine and by a gp120 subunit vaccine.
J Virol 70: 1953-60 (1996)

Mossman SP; Bex F; Berglund P; Arthos J; O'Neil SP; Riley D; Maul DH; Bruck C; Momin P; Burny A; Fultz PN; Mullins JI; Liljestrom P; Hoover EA
Protection against lethal simian immunodeficiency virus SIVsmmPBj14 disease by a recombinant Semliki Forest virus gp160 vaccine and by a gp120 subunit vaccine.
J Virol 70: 1953-60 (1996)

Moutouh L; Corbeil J; Richman DD
Recombination leads to the rapid emergence of HIV-1 dually resistant mutants under selective drug pressure.
Proc Natl Acad Sci U S A 93: 6106-11 (1996)

Mucke L; Abraham CR; Masliah E
Neurotrophic and neuroprotective effects of hAPP in transgenic mice.
Ann N Y Acad Sci 777: 82-8 (1996)

Muriaux D; Fosse P; Paoletti J
A kissing complex together with a stable dimer is involved in the HIV-1Lai RNA dimerization process in vitro.
Biochemistry 35: 5075-82 (1996)

Murphy R; Wente SR
An RNA-export mediator with an essential nuclear export signal.
Nature 383: 357-60 (1996)

Murphy R; Wente SR
An RNA-export mediator with an essential nuclear export signal.
Nature 383: 357-60 (1996)

Musso M; Ghiorzo P; Fiorentini P; Giuffrida R; Ciotti P; Garre C; Ravazzolo R; Bianchi-Scarra G
An upstream positive regulatory element in human GM-CSF promoter is recognized by NF-kappa B/Rel family members.
Biochem Biophys Res Commun 223: 64-72 (1996)

Nador RG; Cesarman E; Chadburn A; Dawson DB; Ansari MQ; Sald J; Knowles DM
Primary effusion lymphoma: a distinct clinicopathologic entity associated with the Kaposi's sarcoma-associated herpes virus.
Blood 88: 645-56 (1996)

Nakano H; Oshima H; Chung W; Williams-Abbott L; Ware CF; Yagita H; Okumura K
TRAF5, an activator of NF-kappaB and putative signal transducer for the lymphotoxin-beta receptor.
J Biol Chem 271: 14661-4 (1996)

Naldini L; Blomer U; Gallay P; Ory D; Mulligan R; Gage FH; Verma IM; Trono D
In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector [see comments]
Science 272: 263-7 (1996)

Narwa R; Roques P; Courpotin C; Parnet-Mathieu F; Boussin F; Roane A; Marce D; Lasfargues G; Dormont D
Characterization of human immunodeficiency virus type 1 p17 matrix protein motifs associated with mother-to-child transmission.
J Virol 70: 4474-83 (1996)

Nashimoto M
Specific cleavage of target RNAs from HIV-1 with 5' half tRNA by mammalian tRNA 3' processing endoribonuclease.
RNA 2: 523-4 (1996)

Nassiri MR; Emerson SG; Devivar RV; Townsend LB; Drach JC; Taichman RS
Comparison of benzimidazole nucleosides and ganciclovir on the in vitro proliferation and colony formation of human bone marrow progenitor cells.
Br J Haematol 93: 273-9 (1996)

Navarro J; Punzon MC; Pizarro A; Fernandez-Cruz E; Fresno M; Munoz-Fernandez MA
Pentoxifylline inhibits acute HIV-1 replication in human T cells by a mechanism not involving inhibition of tumour necrosis factor synthesis or nuclear factor-kappa B activation.
AIDS 10: 469-75 (1996)

Navarro J; Punzon MC; Pizarro A; Fernandez-Cruz E; Fresno M; Munoz-Fernandez MA
Pentoxifylline inhibits acute HIV-1 replication in human T cells by a mechanism not involving inhibition of tumour necrosis factor synthesis or nuclear factor-kappa B activation.
AIDS 10: 469-75 (1996)

Nehete PN; Johnson PC; Schapiro SJ; Arlinghaus RB; Sastry KJ
Cross-reactive T-cell proliferative responses to V3 peptides corresponding to different geographical HIV-1 isolates in HIV-seropositive individuals.
J Clin Immunol 16: 115-24 (1996)

Nelson JS; Giver L; Ellington AD; Letsinger RL
Incorporation of a non-nucleotide bridge into hairpin oligonucleotides capable of high-affinity binding to the Rev protein of HIV-1.
Biochemistry 35: 5339-44 (1996)

Nelson NC; Cheikh AB; Matsuda E; Becker MM
Simultaneous detection of multiple nucleic acid targets in a homogeneous format.
Biochemistry 35: 8429-38 (1996)

Nelson PJ; Ortiz BD; Pattison JM; Krensky AM
Identification of a novel regulatory region critical for expression of the RANTES chemokine in activated T lymphocytes.
J Immunol 157: 1139-48 (1996)

Neurath AR; Jiang S; Strick N; Lin K; Li YY; Debnath AK
Bovine beta-lactoglobulin modified by 3-hydroxyphthalic anhydride blocks the CD4 cell receptor for HIV.
Nat Med 2: 230-4 (1996)

Neurath AR; Jiang S; Strick N; Lin K; Li YY; Debnath AK
Bovine beta-lactoglobulin modified by 3-hydroxyphthalic anhydride blocks the CD4 cell receptor for HIV.
Nat Med 2: 230-4 (1996)

Niikura M; Dornadula G; Zhang H; Mukhtar M; Lingxun D; Khalili K; Bagasra O; Pomerantz RJ
Mechanisms of transcriptional transactivation and restriction of human immunodeficiency virus type I replication in an astrocytic glial cell.
Oncogene 13: 313-22 (1996)

Nikolaev LG
[Identification and isolation of proteins, recognizing the sequence of the human immunodeficiency virus (HIV-1) enhancer]
Mol Biol (Mosk) 30: 714-20 (1996)

Nikolaev LG
[Identification and isolation of proteins, recognizing the sequence of the human immunodeficiency virus (HIV-1) enhancer]
Mol Biol (Mosk) 30: 714-20 (1996)

Nikolakaki E; Simos G; Georgatos SD; Giannakouros T
A nuclear envelope-associated kinase phosphorylates arginine-serine motifs and modulates interactions between the lamin B receptor and other nuclear proteins.
J Biol Chem 271: 8365-72 (1996)

Nilsen BM; Haugan IR; Berg K; Olsen L; Brown PO; Helland DE
Monoclonal antibodies against human immunodeficiency virus type 1 integrase: epitope mapping and differential effects on integrase activities in vitro.
J Virol 70: 1580-7 (1996)

Nissley DV; Garfinkel DJ; Strathern JN
HIV reverse transcription in yeast [letter]
Nature 380: 30 (1996)

Niwa Y; Yano M; Futaki S; Okumura Y; Kido H
T-cell membrane-associated serine protease, tryptase TL2, binds human immunodeficiency virus type 1 gp120 and cleaves the third-variable-domain loop of gp120. Neutralizing antibodies of human immunodeficiency virus type 1 inhibit cleavage of gp120.
Eur J Biochem 237: 64-70 (1996)

Norgard MV; Arndt LL; Akins DR; Curetty LL; Harrich DA; Radolf JD
Activation of human monocytic cells by Treponema pallidum and Borrelia burgdorferi lipoproteins and synthetic lipopeptides proceeds via a pathway distinct from that of lipopolysaccharide but involves the transcriptional activator NF-kappa B.
Infect Immun 64: 3845-52 (1996)

Norgard MV; Arndt LL; Akins DR; Curetty LL; Harrich DA; Radolf JD
Activation of human monocytic cells by Treponema pallidum and Borrelia burgdorferi lipoproteins and synthetic lipopeptides proceeds via a pathway distinct from that of lipopolysaccharide but involves the transcriptional activator NF-kappa B.
Infect Immun 64: 3845-52 (1996)

Norley S; Beer B; Binninger-Schinzel D; Cosma C; Kurth R
Protection from pathogenic SIVmac challenge following short-term infection with a nef-deficient attenuated virus.
Virology 219: 195-205 (1996)

Nottet HS; Persidsky Y; Sasseville VG; Nukuna AN; Bock P; Zhai QH; Sharer LR; McComb RD; Swindells S; Soderland C; Gendelman HE
Mechanisms for the transendothelial migration of HIV-1-infected monocytes into brain.
J Immunol 156: 1284-95 (1996)

Novembre FJ; Lewis MG; Saucier MM; Yalley-Ogunro J; Brennan T; McKinnon K; Bellah S; McClure HM
Deletion of the nef gene abrogates the ability of SIV smmPBj to induce acutely lethal disease in pigtail macaques.
AIDS Res Hum Retroviruses 12: 727-36 (1996)

Novembre FJ; Lewis MG; Saucier MM; Yalley-Ogunro J; Brennan T; McKinnon K; Bellah S; McClure HM
Deletion of the nef gene abrogates the ability of SIV smmPBj to induce acutely lethal disease in pigtail macaques.
AIDS Res Hum Retroviruses 12: 727-36 (1996)

Nowak MA; Bangham CR
Population dynamics of immune responses to persistent viruses.
Science 272: 74-9 (1996)

Nugiel DA; Jacobs K; Kaltenbach RF; Worley T; Patel M; Meyer DT; Jadhav PK; De Lucca GV; Smyser TE; Klabe RM; Bacheler LT; Rayner MM; Seitz SP
Preparation and structure-activity relationship of novel P1/P1'-substituted cyclic urea-based human immunodeficiency virus type-1 protease inhibitors.
J Med Chem 39: 2156-69 (1996)

Nunn MF; Marsh JW
Human immunodeficiency virus type 1 Nef associates with a member of the p21-activated kinase family.
J Virol 70: 6157-61 (1996)

O'Brien WA; Sumner-Smith M; Mao SH; Sadeghi S; Zhao JQ; Chen IS
Anti-human immunodeficiency virus type 1 activity of an oligocationic compound mediated via gp120 V3 interactions.
J Virol 70: 2825-31 (1996)

O'Neil LL; Burkhard MJ; Hoover EA
Frequent perinatal transmission of feline immunodeficiency virus by chronically infected cats.
J Virol 70: 2894-901 (1996)

Ochsenbauer C; Bosch V; Oelze I; Wieland U
Unimpaired function of a naturally occurring C terminally truncated vif gene product of human immunodeficiency virus type 1.
J Gen Virol 77 (Pt 7): 1389-95 (1996)

Olaleye DO; Sheng Z; Howard TM; Rasheed S
Isolation and characterization of a new subtype A variant of human immunodeficiency virus type I from Nigeria.
Trop Med Int Health 1: 97-106 (1996)

Olmsted RA; Slade DE; Kopta LA; Poppe SM; Poel TJ; Newport SW; Rank KB; Biles C; Morge RA; Dueweke TJ; Yagi Y; Romero DL; Thomas RC; Sharma SK; Tarpley WG
(Alkylamino) piperidine bis(heteroaryl)piperizine analogs are potent, broad-spectrum nonnucleoside reverse transcriptase inhibitors of drug-resistant isolates of human immunodeficiency virus type 1 (HIV-1) and select for drug-resistant variants of HIV-1IIIB with reduced replication phenotypes.
J Virol 70: 3698-705 (1996)

Oravecz T; Pall M; Norcross MA
Beta-chemokine inhibition of monocytotropic HIV-1 infection. Interference with a postbinding fusion step.
J Immunol 157: 1329-32 (1996)

Orlando P; Strazzullo G; Carretta F; De Falco M; Grippo P
Inhibition mechanisms of HIV-1, Mo-MuLV and AMV reverse transcriptases by Kelletinin A from Buccinulum corneum.
Experientia 52: 812-7 (1996)

Orlando P; Strazzullo G; Carretta F; De Falco M; Grippo P
Inhibition mechanisms of HIV-1, Mo-MuLV and AMV reverse transcriptases by Kelletinin A from Buccinulum corneum.
Experientia 52: 812-7 (1996)

Orsini MJ; Debouck CM; Webb CL; Lysko PG
Extracellular human immunodeficiency virus type 1 Tat protein promotes aggregation and adhesion of cerebellar neurons.
J Neurosci 16: 2546-52 (1996)

Otsuki T; Kumar S; Ensoli B; Kingma DW; Yano T; Stetler-Stevenson M; Jaffe ES; Raffeld M
Detection of HHV-8/KSHV DNA sequences in AIDS-associated extranodal lymphoid malignancies.
Leukemia 10: 1358-62 (1996)

Otsyula MG; Miller CJ; Marthas ML; Van Rompay KK; Collins JR; Pedersen NC; McChesney MB
Virus-induced immunosuppression is linked to rapidly fatal disease in infant rhesus macaques infected with simian immunodeficiency virus.
Pediatr Res 39: 630-5 (1996)

Otsyula MG; Miller CJ; Marthas ML; Van Rompay KK; Collins JR; Pedersen NC; McChesney MB
Virus-induced immunosuppression is linked to rapidly fatal disease in infant rhesus macaques infected with simian immunodeficiency virus.
Pediatr Res 39: 630-5 (1996)

Otsyula MG; Miller CJ; Marthas ML; Van Rompay KK; Collins JR; Pedersen NC; McChesney MB
Virus-induced immunosuppression is linked to rapidly fatal disease in infant rhesus macaques infected with simian immunodeficiency virus.
Pediatr Res 39: 630-5 (1996)

Otteken A; Moss B
Calreticulin interacts with newly synthesized human immunodeficiency virus type 1 envelope glycoprotein, suggesting a chaperone function similar to that of calnexin.
J Biol Chem 271: 97-103 (1996)

Ouali M; Bouziane M; Ketterle C; Gabarro-Arpa J; Auclair C; Le Bret M
A molecular mechanics and dynamics study of alternate triple-helices involving the integrase-binding site of the HIV-1 virus and oligonucleotides having a 3'-3' internucleotide junction.
J Biomol Struct Dyn 13: 835-53 (1996)

Oyaizu N; McCloskey TW; Than S; Pahwa S
Inhibition of CD4 cross-linking-induced lymphocytes apoptosis by vesnarinone as a novel immunomodulating agent: vesnarinone inhibits Fas expression and apoptosis by blocking cytokine secretion.
Blood 87: 2361-8 (1996)

Paillart JC; Skripkin E; Ehresmann B; Ehresmann C; Marquet R
A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA.
Proc Natl Acad Sci U S A 93: 5572-7 (1996)

Paillart JC; Skripkin E; Ehresmann B; Ehresmann C; Marquet R
The use of chemical modification interference and inverse PCR mutagenesis to identify the dimerization initiation site of HIV-1 genomic RNA.
Pharm Acta Helv 71: 21-8 (1996)

Palaniappan C; Fuentes GM; Rodriguez-Rodriguez L; Fay PJ; Bambara RA
Helix structure and ends of RNA/DNA hybrids direct the cleavage specificity of HIV-1 reverse transcriptase RNase H.
J Biol Chem 271: 2063-70 (1996)

Palmer C; Balfe P; Fox D; May JC; Frederiksson R; Fenyo EM; McKeating JA
Functional characterization of the V1V2 region of human immunodeficiency virus type 1.
Virology 220: 436-49 (1996)

Pampuro SE; Calarota SA; Marquina SA; Rabinovich RD; Libonatti OV
Reactivity of Argentine serum samples against synthetic V3-based HIV-1 peptides [letter]
J Acquir Immune Defic Syndr Hum Retrovirol 12: 527-8 (1996)

Pan Z; Radding W; Zhou T; Hunter E; Mountz J; McDonald JM
Role of calmodulin in HIV-potentiated Fas-mediated apoptosis.
Am J Pathol 149: 903-10 (1996)

Pan Z; Radding W; Zhou T; Hunter E; Mountz J; McDonald JM
Role of calmodulin in HIV-potentiated Fas-mediated apoptosis.
Am J Pathol 149: 903-10 (1996)

Pandey VN; Kaushik N; Rege N; Sarafianos SG; Yadav PN; Modak MJ
Role of methionine 184 of human immunodeficiency virus type-1 reverse transcriptase in the polymerase function and fidelity of DNA synthesis.
Biochemistry 35: 2168-79 (1996)

Pandori MW; Fitch NJ; Craig HM; Richman DD; Spina CA; Guatelli JC
Producer-cell modification of human immunodeficiency virus type 1: Nef is a virion protein.
J Virol 70: 4283-90 (1996)

Panozzo J; Akan E; Griffiths TD; Woloschak GE
The effects of 5-fluorouracil and doxorubicin on expression of human immunodeficiency virus type 1 long terminal repeat.
Cancer Lett 105: 217-23 (1996)

Panozzo J; Panozzo J; Akan E; Libertin C; Woloschak GE
The effects of cisplatin and methotrexate on the expression of human immunodeficiency virus type 1 long terminal repeat.
Leuk Res 20: 309-17 (1996)

Papandreou MJ; Idziorek T; Miquelis R; Fenouillet E
Glycosylation and stability of mature HIV envelope glycoprotein conformation under various conditions.
FEBS Lett 379: 171-6 (1996)

Parent LJ; Wilson CB; Resh MD; Wills JW
Evidence for a second function of the MA sequence in the Rous sarcoma virus Gag protein.
J Virol 70: 1016-26 (1996)

Park IW; Kondo E; Bergeron L; Park J; Sodroski J
Effects of human immunodeficiency virus type 1 infection on programmed cell death in the presence or absence of Bcl-2.
J Acquir Immune Defic Syndr Hum Retrovirol 12: 321-8 (1996)

Park IW; Sodroski J
Targeting a foreign protein into virion particles by fusion with the Vpx protein of simian immunodeficiency virus.
J Acquir Immune Defic Syndr Hum Retrovirol 11: 341-50 (1996)

Parker CE; Papac DI; Trojak SK; Tomer KB
Epitope mapping by mass spectrometry: determination of an epitope on HIV-1 IIIB p26 recognized by a monoclonal antibody.
J Immunol 157: 198-206 (1996)

Patki AH; Lederman MM
HIV-1 Tat protein and its inhibitor Ro 24-7429 inhibit lymphocyte proliferation and induce apoptosis in peripheral blood mononuclear cells from healthy donors.
Cell Immunol 169: 40-6 (1996)

Pau CP; Hu DJ; Spruill C; Schable C; Lackritz E; Kai M; George JR; Rayfield MA; Dondero TJ; Williams AE; Busch MP; Brown AE; McCutchan FE; Schochetman G
Surveillance for human immunodeficiency virus type 1 group O infections in the United States.
Transfusion 36: 398-400 (1996)

Pazhanisamy S; Stuver CM; Cullinan AB; Margolin N; Rao BG; Livingston DJ
Kinetic characterization of human immunodeficiency virus type-1 protease-resistant variants.
J Biol Chem 271: 17979-85 (1996)

Pazin MJ; Sheridan PL; Cannon K; Cao Z; Keck JG; Kadonaga JT; Jones KA
NF-kappa B-mediated chromatin reconfiguration and transcriptional activation of the HIV-1 enhancer in vitro.
Genes Dev 10: 37-49 (1996)

Pellegrin I; Legrand E; Neau D; Bonot P; Masquelier B; Pellegrin JL; Ragnaud JM; Bernard N; Fleury HJ
Kinetics of appearance of neutralizing antibodies in 12 patients with primary or recent HIV-1 infection and relationship with plasma and cellular viral loads.
J Acquir Immune Defic Syndr Hum Retrovirol 11: 438-47 (1996)

Pemberton IK; Buckle M; Buc H
The metal ion-induced cooperative binding of HIV-1 integrase to DNA exhibits a marked preference for Mn(II) rather than Mg(II).
J Biol Chem 271: 1498-506 (1996)

Peng H; Callison D; Li P; Burrell C
Long-term protection against HIV-1 infection conferred by tat or rev antisense RNA was affected by the design of the retroviral vector.
Virology 220: 377-89 (1996)

Pengue G; Lania L
Kruppel-associated box-mediated repression of RNA polymerase II promoters is influenced by the arrangement of basal promoter elements.
Proc Natl Acad Sci U S A 93: 1015-20 (1996)

Penny MA; Thomas SJ; Douglas NW; Ranjbar S; Holmes H; Daniels RS
env gene sequences of primary HIV type 1 isolates of subtypes B, C, D, E, and F obtained from the World Health Organization Network for HIV Isolation and Characterization.
AIDS Res Hum Retroviruses 12: 741-7 (1996)

Penny MA; Thomas SJ; Douglas NW; Ranjbar S; Holmes H; Daniels RS
env gene sequences of primary HIV type 1 isolates of subtypes B, C, D, E, and F obtained from the World Health Organization Network for HIV Isolation and Characterization.
AIDS Res Hum Retroviruses 12: 741-7 (1996)

Pezo V; Martinez MA; Wain-Hobson S
Fate of direct and inverted repeats in the RNA hypermutagenesis reaction.
Nucleic Acids Res 24: 253-6 (1996)

Pham-Kanter GB; Steinberg MH; Ballard RC
Sexually transmitted diseases in South Africa.
Genitourin Med 72: 160-71 (1996)

Pilkington GR; Duan L; Zhu M; Keil W; Pomerantz RJ
Recombinant human Fab antibody fragments to HIV-1 Rev and Tat regulatory proteins: direct selection from a combinatorial phage display library.
Mol Immunol 33: 439-50 (1996)

Pillay D; Geddes AM
Antiviral drug resistance [editorial]
BMJ 313: 503-4 (1996)

Pillay D; Geddes AM
Antiviral drug resistance [editorial]
BMJ 313: 503-4 (1996)

Piller SC; Ewart GD; Premkumar A; Cox GB; Gage PW
Vpr protein of human immunodeficiency virus type 1 forms cation-selective channels in planar lipid bilayers.
Proc Natl Acad Sci U S A 93: 111-5 (1996)

Planelles V; Jowett JB; Li QX; Xie Y; Hahn B; Chen IS
Vpr-induced cell cycle arrest is conserved among primate lentiviruses.
J Virol 70: 2516-24 (1996)

Planelles V; Jowett JB; Li QX; Xie Y; Hahn B; Chen IS
Vpr-induced cell cycle arrest is conserved among primate lentiviruses.
J Virol 70: 2516-24 (1996)

Poccia F; Boullier S; Lecoeur H; Cochet M; Poquet Y; Colizzi V; Fournie JJ; Gougeon ML
Peripheral V gamma 9/V delta 2 T cell deletion and anergy to nonpeptidic mycobacterial antigens in asymptomatic HIV-1-infected persons.
J Immunol 157: 449-61 (1996)

Poignard P; Fouts T; Naniche D; Moore JP; Sattentau QJ
Neutralizing antibodies to human immunodeficiency virus type-1 gp120 induce envelope glycoprotein subunit dissociation.
J Exp Med 183: 473-84 (1996)

Pokholok DK; Gudima SO; Memelova LV; Esipov DS; Rechinskii VO; Kochetkov SN
[Stability of human immunodeficiency virus to azidothymidine. II. Kinetic characteristics of "AZT-resistant" mutant forms of reverse transcriptase]
Biokhimiia 61: 142-51 (1996)

Pop MP
In vitro analysis of the HIV-1 second strand-transfer reaction.
Biochim Biophys Acta 1307: 193-204 (1996)

Pop MP; Biebricher CK
Kinetic analysis of pausing and fidelity of human immunodeficiency virus type 1 reverse transcription.
Biochemistry 35: 5054-62 (1996)

Porter DC; Melsen LR; Compans RW; Morrow CD
Release of virus-like particles from cells infected with poliovirus replicons which express human immunodeficiency virus type 1 Gag.
J Virol 70: 2643-9 (1996)

Powell PD
Can evolutionary principles help resolve the acquired immune deficiency syndrome crisis?
Med Hypotheses 46: 130-4 (1996)

Pratt RD; Nichols S; McKinney N; Kwok S; Dankner WM; Spector SA
Virologic markers of human immunodeficiency virus type 1 in cerebrospinal fluid of infected children.
J Infect Dis 174: 288-93 (1996)

Prochaska HJ; Fernandes CL; Pantoja RM; Chavan SJ
Inhibition of human immunodeficiency virus type 1 long terminal repeat-driven transcription by an in vivo metabolite of oltipraz: implications for antiretroviral therapy.
Biochem Biophys Res Commun 221: 548-53 (1996)

Puttaraju M; Been MD
Circularizing ribozymes and decoy-competitors by autocatalytic splicing in vitro and in vivo.
SAAS Bull Biochem Biotechnol 9: 77-82 (1996)

Qavi HB; Xu B; Green MT; Lusso P; Pearson G; Ablashi DV
Morphological and ultrastructural changes induced in corneal epithelial cells by HIV-1 and HHV-6 in vitro.
Curr Eye Res 15: 597-604 (1996)

Qavi HB; Xu B; Green MT; Lusso P; Pearson G; Ablashi DV
Morphological and ultrastructural changes induced in corneal epithelial cells by HIV-1 and HHV-6 in vitro.
Curr Eye Res 15: 597-604 (1996)

Quereda C; Polanco AM; Giner C; Sanchez-Sousa A; Pereira E; Navas E; Fortun J; Guerrero A; Baquero F
Correlation between in vitro resistance to fluconazole and clinical outcome of oropharyngeal candidiasis in HIV-infected patients.
Eur J Clin Microbiol Infect Dis 15: 30-7 (1996)

Quillent C; Borman AM; Paulous S; Dauguet C; Clavel F
Extensive regions of pol are required for efficient human immunodeficiency virus polyprotein processing and particle maturation.
Virology 219: 29-36 (1996)

Radding W; Pan ZQ; Hunter E; Johnston P; Williams JP; McDonald JM
Expression of HIV-1 envelope glycoprotein alters cellular calmodulin.
Biochem Biophys Res Commun 218: 192-7 (1996)

Radiuk SN; Matsevich GR; Ryzhov KA; Lariukova TA
[Nonisotopic variant of quantitative analysis using polymerase chain reaction for diagnosing HIV infection]
Vopr Virusol 41: 44-5 (1996)

Ragni MV; Koch WC; Jordan JA
Parvovirus B19 infection in patients with hemophilia.
Transfusion 36: 238-41 (1996)

Raillard SA; Joyce GF
Targeting sites within HIV-1 cDNA with a DNA-cleaving ribozyme.
Biochemistry 35: 11693-701 (1996)

Raillard SA; Joyce GF
Targeting sites within HIV-1 cDNA with a DNA-cleaving ribozyme.
Biochemistry 35: 11693-701 (1996)

Raja NU; Jabbar MA
The human immunodeficiency virus type 1 Vpu protein tethered to the CD4 extracellular domain is localized to the plasma membrane and is biologically active in the secretory pathway of mammalian cells: implications for the mechanisms of Vpu function.
Virology 220: 141-51 (1996)

Ramazzotti E; Vignoli M; Re MC; Furlini G; La Placa M
Enhanced nuclear factor-kappa B activation induced by tumour necrosis factor-alpha in stably tat-transfected cells is associated with the presence of cell-surface-bound Tat protein.
AIDS 10: 455-61 (1996)

Ramazzotti E; Vignoli M; Re MC; Furlini G; La Placa M
Enhanced nuclear factor-kappa B activation induced by tumour necrosis factor-alpha in stably tat-transfected cells is associated with the presence of cell-surface-bound Tat protein.
AIDS 10: 455-61 (1996)

Rasty S; Thatikunta P; Gordon J; Khalili K; Amini S; Glorioso JC
Human immunodeficiency virus tat gene transfer to the murine central nervous system using a replication-defective herpes simplex virus vector stimulates transforming growth factor beta 1 gene expression.
Proc Natl Acad Sci U S A 93: 6073-8 (1996)

Ratto S; Sitz KV; Scherer AM; Loomis LD; Cox JH; Redfield RR; Birx DL
CD4+ T-lymphocyte lines developed from HIV-1-seropositive patients recognize different epitopes within the V3 loop.
J Acquir Immune Defic Syndr Hum Retrovirol 11: 128-36 (1996)

Reddy DV; Jagannadh B; Kunwar AC
NMR study of dideoxynucleotides with anti-human immunodeficiency virus (HIV) activity.
J Biochem Biophys Methods 31: 113-21 (1996)

Reddy RT; Achim CL; Sirko DA; Tehranchi S; Kraus FG; Wong-Staal F; Wiley CA
Sequence analysis of the V3 loop in brain and spleen of patients with HIV encephalitis.
AIDS Res Hum Retroviruses 12: 477-82 (1996)

Reich SH; Melnick M; Pino MJ; Fuhry MA; Trippe AJ; Appelt K; Davies JF 2nd; Wu BW; Musick L
Structure-based design and synthesis of substituted 2-butanols as nonpeptidic inhibitors of HIV protease: secondary amide series.
J Med Chem 39: 2781-94 (1996)

Reifsnyder C; Lowell J; Clarke A; Pillus L
Yeast SAS silencing genes and human genes associated with AML and HIV-1 Tat interactions are homologous with acetyltransferases [see comments]
Nat Genet 14: 42-9 (1996)

Reifsnyder C; Lowell J; Clarke A; Pillus L
Yeast SAS silencing genes and human genes associated with AML and HIV-1 Tat interactions are homologous with acetyltransferases [see comments]
Nat Genet 14: 42-9 (1996)

Reimann KA; Li JT; Voss G; Lekutis C; Tenner-Racz K; Racz P; Lin W; Montefiori DC; Lee-Parritz DE; Lu Y; Collman RG; Sodroski J; Letvin NL
An env gene derived from a primary human immunodeficiency virus type 1 isolate confers high in vivo replicative capacity to a chimeric simian/human immunodeficiency virus in rhesus monkeys.
J Virol 70: 3198-206 (1996)

Reimann KA; Li JT; Voss G; Lekutis C; Tenner-Racz K; Racz P; Lin W; Montefiori DC; Lee-Parritz DE; Lu Y; Collman RG; Sodroski J; Letvin NL
An env gene derived from a primary human immunodeficiency virus type 1 isolate confers high in vivo replicative capacity to a chimeric simian/human immunodeficiency virus in rhesus monkeys.
J Virol 70: 3198-206 (1996)

Renne R; Zhong W; Herndier B; McGrath M; Abbey N; Kedes D; Ganem D
Lytic growth of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in culture.
Nat Med 2: 342-6 (1996)

Rey O; Canon J; Krogstad P
HIV-1 Gag protein associates with F-actin present in microfilaments.
Virology 220: 530-4 (1996)

Rice P; Craigie R; Davies DR
Retroviral integrases and their cousins.
Curr Opin Struct Biol 6: 76-83 (1996)

Richman DD
Antiretroviral drug resistance: mechanisms, pathogenesis, clinical significance.
Adv Exp Med Biol 394: 383-95 (1996)

Richman DD
The implications of drug resistance for strategies of combination antiviral chemotherapy.
Antiviral Res 29: 31-3 (1996)

Richman DD
Antiretroviral drug resistance: mechanisms, pathogenesis, clinical significance.
Adv Exp Med Biol 394: 383-95 (1996)

Richman DD
The implications of drug resistance for strategies of combination antiviral chemotherapy.
Antiviral Res 29: 31-3 (1996)

Riddell SR; Elliott M; Lewinsohn DA; Gilbert MJ; Wilson L; Manley SA; Lupton SD; Overell RW; Reynolds TC; Corey L; Greenberg PD
T-cell mediated rejection of gene-modified HIV-specific cytotoxic T lymphocytes in HIV-infected patients [see comments]
Nat Med 2: 216-23 (1996)

Ridky TW; Bizub-Bender D; Cameron CE; Weber IT; Wlodawer A; Copeland T; Skalka AM; Leis J
Programming the Rous sarcoma virus protease to cleave new substrate sequences.
J Biol Chem 271: 10538-44 (1996)

Ridky TW; Cameron CE; Cameron J; Leis J; Copeland T; Wlodawer A; Weber IT; Harrison RW
Human immunodeficiency virus, type 1 protease substrate specificity is limited by interactions between substrate amino acids bound in adjacent enzyme subsites.
J Biol Chem 271: 4709-17 (1996)

Riggs NL; Guatelli JC
Production and characterization of high-titer stocks of rev-defective HIV-1.
Virology 217: 602-6 (1996)

Rill RL; Hecker KH
Sequence-specific actinomycin D binding to single-stranded DNA inhibits HIV reverse transcriptase and other polymerases.
Biochemistry 35: 3525-33 (1996)

Ritter GD Jr; Yamshchikov G; Cohen SJ; Mulligan MJ
Human immunodeficiency virus type 2 glycoprotein enhancement of particle budding: role of the cytoplasmic domain.
J Virol 70: 2669-73 (1996)

Roberts CG; Meister GE; Jesdale BM; Lieberman J; Berzofsky JA; De Groot AS
Prediction of HIV peptide epitopes by a novel algorithm.
AIDS Res Hum Retroviruses 12: 593-610 (1996)

Roberts CG; Meister GE; Jesdale BM; Lieberman J; Berzofsky JA; De Groot AS
Prediction of HIV peptide epitopes by a novel algorithm.
AIDS Res Hum Retroviruses 12: 593-610 (1996)

Robey FA; Harris-Kelson T; Robert-Guroff M; Batinic D; Ivanov B; Lewis MS; Roller PP
A synthetic conformational epitope from the C4 domain of HIV Gp120 that binds CD4.
J Biol Chem 271: 17990-5 (1996)

Robinson CA; Rose NC
Tuberculosis: current implications and management in obstetrics.
Obstet Gynecol Surv 51: 115-24 (1996)

Robinson WE Jr; Reinecke MG; Abdel-Malek S; Jia Q; Chow SA
Inhibitors of HIV-1 receptor that inhibit HIV integrase.
Proc Natl Acad Sci U S A 93: 6326-31 (1996)

Rodriguez-Barradas MC; Groover JE; Lacke CE; Gump DW; Lahart CJ; Pandey JP; Musher DM
IgG antibody to pneumococcal capsular polysaccharide in human immunodeficiency virus-infected subjects: persistence of antibody in responders, revaccination in nonresponders, and relationship of immunoglobulin allotype to response.
J Infect Dis 173: 1347-53 (1996)

Romano MF; Lamberti A; Petrella A; Bisogni R; Tassone PF; Formisano S; Venuta S; Turco MC
IL-10 inhibits nuclear factor-kappa B/Rel nuclear activity in CD3-stimulated human peripheral T lymphocytes.
J Immunol 156: 2119-23 (1996)

Rose RE; Gong YF; Greytok JA; Bechtold CM; Terry BJ; Robinson BS; Alam M; Colonno RJ; Lin PF
Human immunodeficiency virus type 1 viral background plays a major role in development of resistance to protease inhibitors.
Proc Natl Acad Sci U S A 93: 1648-53 (1996)

Rosenzweig M; Marks DF; Zhu H; Hempel D; Mansfield KG; Sehgal PK; Kalams S; Scadden DT; Johnson RP
In vitro T lymphopoiesis of human and rhesus CD34+ progenitor cells.
Blood 87: 4040-8 (1996)

Rossi F; Gallina A; Milanesi G
Nef-CD4 physical interaction sensed with the yeast two-hybrid system.
Virology 217: 397-403 (1996)

Rounseville MP; Lin HC; Agbottah E; Shukla RR; Rabson AB; Kumar A
Inhibition of HIV-1 replication in viral mutants with altered TAR RNA stem structures.
Virology 216: 411-7 (1996)

Rozera C; Carattoli A; De Marco A; Amici C; Giorgi C; Santoro MG
Inhibition of HIV-1 replication by cyclopentenone prostaglandins in acutely infected human cells. Evidence for a transcriptional block.
J Clin Invest 97: 1795-803 (1996)

Ruprecht RM; Baba TW; Liska V
Attenuated HIV vaccine: caveats [letter]
Science 271: 1790-2 (1996)

Ruprecht RM; Baba TW; Liska V
Attenuated HIV vaccine: caveats [letter]
Science 271: 1790-2 (1996)

Said W; Chien K; Takeuchi S; Tasaka T; Asou H; Cho SK; de Vos S; Cesarman E; Knowles DM; Koeffler HP
Kaposi's sarcoma-associated herpesvirus (KSHV or HHV8) in primary effusion lymphoma: ultrastructural demonstration of herpesvirus in lymphoma cells.
Blood 87: 4937-43 (1996)

Samson M; Libert F; Doranz BJ; Rucker J; Liesnard C; Farber CM; Saragosti S; Lapoumeroulie C; Cognaux J; Forceille C; Muyldermans G; Verhofstede C; Burtonboy G; Georges M; Imai T; Rana S; Yi Y; Smyth RJ; Collman RG; Doms RW; Vassart G; Parmentier M
Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene [see comments]
Nature 382: 722-5 (1996)

Sandstrom PA; Folks TM
New strategies for treating AIDS.
Bioessays 18: 343-6 (1996)

Sandstrom PA; Pardi D; Goldsmith CS; Chengying D; Diamond AM; Folks TM
bc1-2 expression facilitates human immunodeficiency virus type-1 mediated cytopathic effects during acute spreading infections.
J Virol 70: 4617-22 (1996)

Santamaria-Fries M; Fajardo LF; Sogin ML; Olson PD; Relman DA
Lethal infection by a previously unrecognised metazoan parasite.
Lancet 347: 1797-801 (1996)

Sasseville VG; Du Z; Chalifoux LV; Pauley DR; Young HL; Sehgal PK; Desrosiers RC; Lackner AA
Induction of lymphocyte proliferation and severe gastrointestinal disease in macaques by a nef gene variant SIVmac239.
Am J Pathol 149: 163-76 (1996)

Sastry KJ; Marin MC; Nehete PN; McConnell K; el-Naggar AK; McDonnell TJ
Expression of human immunodeficiency virus type I tat results in down-regulation of bcl-2 and induction of apoptosis in hematopoietic cells.
Oncogene 13: 487-93 (1996)

Sato A; Yoshimoto J; Isaka Y; Miki S; Suyama A; Adachi A; Hayami M; Fujiwara T; Yoshie O
Evidence for direct association of Vpr and matrix protein p17 within the HIV-1 virion.
Virology 220: 208-12 (1996)

Sauter MM; Pelchen-Matthews A; Bron R; Marsh M; LaBranche CC; Vance PJ; Romano J; Haggarty BS; Hart TK; Lee WM; Hoxie JA
An internalization signal in the simian immunodeficiency virus transmembrane protein cytoplasmic domain modulates expression of envelope glycoproteins on the cell surface.
J Cell Biol 132: 795-811 (1996)

Saxena SK; Gravell M; Wu YN; Mikulski SM; Shogen K; Ardelt W; Youle RJ
Inhibition of HIV-1 production and selective degradation of viral RNA by an amphibian ribonuclease.
J Biol Chem 271: 20783-8 (1996)

Saxena SK; Gravell M; Wu YN; Mikulski SM; Shogen K; Ardelt W; Youle RJ
Inhibition of HIV-1 production and selective degradation of viral RNA by an amphibian ribonuclease.
J Biol Chem 271: 20783-8 (1996)

Sayre KR; Dodd RY; Tegtmeier G; Layug L; Alexander SS; Busch MP
False-positive human immunodeficiency virus type 1 western blot tests in noninfected blood donors.
Transfusion 36: 45-52 (1996)

Schaeffer F; Rimsky S; Spassky A
DNA-stacking interactions determine the sequence specificity of the deoxyribonuclease activity of 1,10-phenanthroline-copper ion.
J Mol Biol 260: 523-39 (1996)

Schapiro JM; Winters MA; Stewart F; Efron B; Norris J; Kozal MJ; Merigan TC
The effect of high-dose saquinavir on viral load and CD4+ T-cell counts in HIV-infected patients [see comments]
Ann Intern Med 124: 1039-50 (1996)

Schmalzbauer E; Strack B; Dannull J; Guehmann S; Moelling K
Mutations of basic amino acids of NCp7 of human immunodeficiency virus type 1 affect RNA binding in vitro.
J Virol 70: 771-7 (1996)

Schmidtmayerova H; Nottet HS; Nuovo G; Raabe T; Flanagan CR; Dubrovsky L; Gendelman HE; Cerami A; Bukrinsky M; Sherry B
Human immunodeficiency virus type 1 infection alters chemokine beta peptide expression in human monocytes: implications for recruitment of leukocytes into brain and lymph nodes.
Proc Natl Acad Sci U S A 93: 700-4 (1996)

Schmoeckel C
[Kaposi sarcoma in Caucasian women. Clinical, chemical laboratory and endocrinologic examination of 8 women with HIV associated or classical Kaposi sarcoma (letter)]
Hautarzt 47: 394 (1996)

Schonning K; Jansson B; Olofsson S; Hansen JE
Rapid selection for an N-linked oligosaccharide by monoclonal antibodies directed against the V3 loop of human immunodeficiency virus type 1.
J Gen Virol 77 (Pt 4): 753-8 (1996)

Schonning K; Jansson B; Olofsson S; Nielsen JO; Hansen JS
Resistance to V3-directed neutralization caused by an N-linked oligosaccharide depends on the quaternary structure of the HIV-1 envelope oligomer.
Virology 218: 134-40 (1996)

Schooley RT; Campbell TB; Kuritzkes DR; Blaschke T; Stein DS; Rosandich ME; Phair J; Pottage JC; Messari F; Collier A; Kahn J
Phase 1 study of combination therapy with L-697,661 and zidovudine. The ACTG 184 Protocol Team.
J Acquir Immune Defic Syndr Hum Retrovirol 12: 363-70 (1996)

Schooley RT; Ramirez-Ronda C; Lange JM; Cooper DA; Lavelle J; Lefkowitz L; Moore M; Larder BA; St. Clair M; Mulder JW; McKinnis R; Pennington KN; Harrigan PR; Kinghorn I; Steel H; Rooney JF
Virologic and immunologic benefits of initial combination therapy with zidovudine and zalcitabine or didanosine compared with zidovudine monotherapy. Wellcome Resistance Study Collaborative Group.
J Infect Dis 173: 1354-66 (1996)

Schubert U; Bour S; Ferrer-Montiel AV; Montal M; Maldarell F; Strebel K
The two biological activities of human immunodeficiency virus type 1 Vpu protein involve two separable structural domains.
J Virol 70: 809-19 (1996)

Schwartz O; Marechal V; Le Gall S; Lemonnier F; Heard JM
Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein.
Nat Med 2: 338-42 (1996)

Schwille P; Oehlenschlager F; Walter NG
Quantitative hybridization kinetics of DNA probes to RNA in solution followed by diffusional fluorescence correlation analysis.
Biochemistry 35: 10182-93 (1996)

Segall H; Lubin I; Marcus H; Canaan A; Reisner Y
Generation of primary antigen-specific human cytotoxic T lymphocytes in human/mouse radiation chimera.
Blood 88: 721-30 (1996)

Sehgal S; Pasricha N; Jamil S
Serotype analysis of Indian patients with HIV infection.
Trop Med Int Health 1: 199-204 (1996)

Sei S; Akiyoshi H; Bernard J; Venzon DJ; Fox CH; Schwartzentruber DJ; Anderson BD; Kopp JB; Mueller BU; Pizzo PA
Dynamics of virus versus host interaction in children with human immunodeficiency virus type 1 infection.
J Infect Dis 173: 1485-90 (1996)

SenGupta DJ; Zhang B; Kraemer B; Pochart P; Fields S; Wickens M
A three-hybrid system to detect RNA-protein interactions in vivo.
Proc Natl Acad Sci U S A 93: 8496-501 (1996)

Setterquist RA; Smith GK
Ready to use agarose encapsulated PCR reagents.
Nucleic Acids Res 24: 1580-1 (1996)

Shafer RW; Edlin BR
Tuberculosis in patients infected with human immunodeficiency virus: perspective on the past decade.
Clin Infect Dis 22: 683-704 (1996)

Shafer RW; Edlin BR
Tuberculosis in patients infected with human immunodeficiency virus: perspective on the past decade.
Clin Infect Dis 22: 683-704 (1996)

Shafer RW; Winters MA; Jellinger RM; Merigan TC
Zidovudine resistance reverse transcriptase mutations during didanosine monotherapy [letter]
J Infect Dis 174: 448-9 (1996)

Shah K; Davis C; Wilson J; Parekh B
Chimeric synthetic peptides as antigens for detection of antibodies to HIV-1 and HIV-2.
East Afr Med J 73: 63-6 (1996)

Shah K; Neenhold H; Wang Z; Rana TM
Incorporation of an artificial protease and nuclease at the HIV-1 Tat binding site of trans-activation responsive RNA.
Bioconjug Chem 7: 283-9 (1996)

Shah K; Neenhold H; Wang Z; Rana TM
Incorporation of an artificial protease and nuclease at the HIV-1 Tat binding site of trans-activation responsive RNA.
Bioconjug Chem 7: 283-9 (1996)

Shah SM; Shapshak P; Rivers JE; Stewart RV; Weatherby NL; Xin KQ; Page JB; Chitwood DD; Mash DC; Vlahov D; McCoy CB
Detection of HIV-1 DNA in needle/syringes, paraphernalia, and washes from shooting galleries in Miami: a preliminary laboratory report.
J Acquir Immune Defic Syndr Hum Retrovirol 11: 301-6 (1996)

Shaheen F; Duan L; Zhu M; Bagasra O; Pomerantz RJ
Targeting human immunodeficiency virus type 1 reverse transcriptase by intracellular expression of single-chain variable fragments to inhibit early stages of the viral life cycle.
J Virol 70: 3392-400 (1996)

Sham HL; Zhao C; Marsh KC; Betebenner DA; Lin S; Rosenbrook W Jr; Herrin T; Li L; Madigan D; Vasavanonda S; Molla A; Saldivar A; McDonald E; Wideburg NE; Kempf D; Norbeck DW; Plattner JJ
Novel azacyclic ureas that are potent inhibitors of HIV-1 protease.
Biochem Biophys Res Commun 225: 436-40 (1996)

Sham HL; Zhao C; Marsh KC; Betebenner DA; Lin S; Rosenbrook W Jr; Herrin T; Li L; Madigan D; Vasavanonda S; Molla A; Saldivar A; McDonald E; Wideburg NE; Kempf D; Norbeck DW; Plattner JJ
Novel azacyclic ureas that are potent inhibitors of HIV-1 protease.
Biochem Biophys Res Commun 225: 436-40 (1996)

Sham HL; Zhao C; Stewart KD; Betebenner DA; Lin S; Park CH; Kong XP; Rosenbrook W Jr; Herrin T; Madigan D; Vasavanonda S; Lyons N; Molla A; Saldivar A; Marsh KC; McDonald E; Wideburg NE; Denissen JF; Robins T; Kempf DJ; Plattner JJ; Norbeck DW
A novel, picomolar inhibitor of human immunodeficiency virus type 1 protease.
J Med Chem 39: 392-7 (1996)

Sharer LR; Saito Y; Da Cunha A; Ung PC; Gelbard HA; Epstein LG; Blumberg BM
In situ amplification and detection of HIV-1 DNA in fixed pediatric AIDS brain tissue.
Hum Pathol 27: 614-7 (1996)

Sharma V; Xu M; Ritter LM; Wilkie NM
HIV-1 tat induces the expression of a new hematopoietic cell-specific transcription factor and downregulates MIP-1 alpha gene expression in activated T-cells.
Biochem Biophys Res Commun 223: 526-33 (1996)

Sheehy N; Desselberger U; Whitwell H; Ball JK
Concurrent evolution of regions of the envelope and polymerase genes of human immunodeficiency virus type 1 during zidovudine (AZT) therapy.
J Gen Virol 77 (Pt 5): 1071-81 (1996)

Sherman KE; Andreatta C; O'Brien J; Gutierrez A; Harris R
Hepatitis C in human immunodeficiency virus-coinfected patients: increased variability in the hypervariable envelope coding domain.
Hepatology 23: 688-94 (1996)

Shi XP; Yin KC; Zimolo ZA; Stern AM; Waxman L
The subcellular distribution of eukaryotic translation initiation factor, eIF-5A, in cultured cells.
Exp Cell Res 225: 348-56 (1996)

Shibata R; Seimon C; Cho MW; Arthur LO; Nigida SM Jr; Matthews T; Sawyer LA; Schultz A; Murthy KK; Israel Z; Javadian A; Frost P; Kennedy RC; Lane HC; Martin MA
Resistance of previously infected chimpanzees to successive challenges with a heterologous intraclade B strain of human immunodeficiency virus type 1.
J Virol 70: 4361-9 (1996)

Shirai M; Chen M; Arichi T; Masaki T; Nishioka M; Newman M; Nakazawa T; Feinstone SM; Berzofsky JA
Use of intrinsic and extrinsic helper epitopes for in vivo induction of anti-hepatitis C virus cytotoxic T lymphocytes (CTL) with CTL epitope peptide vaccines.
J Infect Dis 173: 24-31 (1996)

Shugars DC; Wild CT; Greenwell TK; Matthews TJ
Biophysical characterization of recombinant proteins expressing the leucine zipper-like domain of the human immunodeficiency virus type 1 transmembrane protein gp41.
J Virol 70: 2982-91 (1996)

Silva AM; Cachau RE; Sham HL; Erickson JW
Inhibition and catalytic mechanism of HIV-1 aspartic protease.
J Mol Biol 255: 321-46 (1996)

Silvestri R; Artico M; Pagnozzi E; Stefancich G; Massa S; La Colla P; Loi AG; Spiga MG; Corrias S; Lichino D
Synthesis and anti-HIV activity of 10,11-dihydropyrrolo [1,2-b][1,2,5]benzothiadiazepine-11-acetic acid 5,5-dioxide derivatives and related compounds.
Farmaco 51: 425-30 (1996)

Silvestri R; Artico M; Pagnozzi E; Stefancich G; Massa S; La Colla P; Loi AG; Spiga MG; Corrias S; Lichino D
Synthesis and anti-HIV activity of 10,11-dihydropyrrolo [1,2-b][1,2,5]benzothiadiazepine-11-acetic acid 5,5-dioxide derivatives and related compounds.
Farmaco 51: 425-30 (1996)

Simon HU; Yousefi S; Dommann-Scherrer CC; Zimmermann DR; Bauer S; Barandun J; Blaser K
Expansion of cytokine-producing CD4-CD8- T cells associated with abnormal Fas expression and hypereosinophilia [see comments]
J Exp Med 183: 1071-82 (1996)

Simons FH; Rutjes SA; van Venrooij WJ; Pruijn GJ
The interactions with Ro60 and La differentially affect nuclear export of hY1 RNA.
RNA 2: 264-73 (1996)

Sirko DA; Ehrlich GD
Genotypic and phenotypic characterization of a neutralization-resistant breakthrough population of HIV-1.
Virology 218: 238-42 (1996)

Sjolander S; Bolmstedt A; Akerblom L; Horal P; Olofsson S; Morein B; Sjolander A
N-linked glycans in the CD4-binding domain of human immunodeficiency virus type 1 envelope glycoprotein gp160 are essential for the in vivo priming of T cells recognizing an epitope located in their vicinity.
Virology 215: 124-33 (1996)

Sjolander S; Hansen JE; Lovgren-Bengtsson K; Akerblom L; Morein B
Induction of homologous virus neutralizing antibodies in guinea-pigs immunized with two human immunodeficiency virus type 1 glycoprotein gp120-iscom preparations. A comparison with other adjuvant systems.
Vaccine 14: 344-52 (1996)

Sjolander S; Hansen JE; Lovgren-Bengtsson K; Akerblom L; Morein B
Induction of homologous virus neutralizing antibodies in guinea-pigs immunized with two human immunodeficiency virus type 1 glycoprotein gp120-iscom preparations. A comparison with other adjuvant systems.
Vaccine 14: 344-52 (1996)

Skripkin E; Isel C; Marquet R; Ehresmann B; Ehresmann C
Psoralen crosslinking between human immunodeficiency virus type 1 RNA and primer tRNA3(Lys).
Nucleic Acids Res 24: 509-14 (1996)

Smith BL; Krushelnycky BW; Mochly-Rosen D; Berg P
The HIV nef protein associates with protein kinase C theta.
J Biol Chem 271: 16753-7 (1996)

Smith KJ; Skelton HG; Hilyard EJ; Hadfield T; Moeller RS; Tuur S; Decker C; Wagner KF; Angritt P
Bacillus piliformis infection (Tyzzer's disease) in a patient infected with HIV-1: confirmation with 16S ribosomal RNA sequence analysis.
J Am Acad Dermatol 34: 343-8 (1996)

Smithgall MD; Wong JG; Critchett KE; Haffar OK
IL-7 up-regulates HIV-1 replication in naturally infected peripheral blood mononuclear cells.
J Immunol 156: 2324-30 (1996)

Sohn MJ; Lee ME; Park HS; Nham SU; Lee YI
Overexpression and purification of human immunodeficiency virus type 1 env derived epitopes in Escherichia coli.
J Biotechnol 45: 211-6 (1996)

Sohn MJ; Lee ME; Park HS; Nham SU; Lee YI
Overexpression and purification of human immunodeficiency virus type 1 env derived epitopes in Escherichia coli.
J Biotechnol 45: 211-6 (1996)

Soler F; Poujade C; Evers M; Carry JC; Henin Y; Bousseau A; Huet T; Pauwels R; De Clercq E; Mayaux JF; Le Pecq JB; Dereu N
Betulinic acid derivatives: a new class of specific inhibitors of human immunodeficiency virus type 1 entry.
J Med Chem 39: 1069-83 (1996)

Songok EM; Tukei PM; Mulaa FJ
Serological investigation of HIV-1 variant subtype strains in transmission in Nairobi.
East Afr Med J 73: 88-90 (1996)

Sonza S; Maerz A; Deacon N; Meanger J; Mills J; Crowe S
Human immunodeficiency virus type 1 replication is blocked prior to reverse transcription and integration in freshly isolated peripheral blood monocytes.
J Virol 70: 3863-9 (1996)

Spira AI; Marx PA; Patterson BK; Mahoney J; Koup RA; Wolinsky SM; Ho DD
Cellular targets of infection and route of viral dissemination after an intravaginal inoculation of simian immunodeficiency virus into rhesus macaques.
J Exp Med 183: 215-25 (1996)

Spira AI; Marx PA; Patterson BK; Mahoney J; Koup RA; Wolinsky SM; Ho DD
Cellular targets of infection and route of viral dissemination after an intravaginal inoculation of simian immunodeficiency virus into rhesus macaques.
J Exp Med 183: 215-25 (1996)

St. Clair MH; Millard J; Rooney J; Tisdale M; Parry N; Sadler BM; Blum MR; Painter G
In vitro antiviral activity of 141W94 (VX-478) in combination with other antiretroviral agents.
Antiviral Res 29: 53-6 (1996)

St. Clair MH; Millard J; Rooney J; Tisdale M; Parry N; Sadler BM; Blum MR; Painter G
In vitro antiviral activity of 141W94 (VX-478) in combination with other antiretroviral agents.
Antiviral Res 29: 53-6 (1996)

Staats HF; Nichols WG; Palker TJ
Mucosal immunity to HIV-1: systemic and vaginal antibody responses after intranasal immunization with the HIV-1 C4/V3 peptide T1SP10 MN(A).
J Immunol 157: 462-72 (1996)

Stanfield-Oakley SA; Griffith JD
Nucleosomal arrangement of HIV-1 DNA: maps generated from an integrated genome and an EBV-based episomal model.
J Mol Biol 256: 503-16 (1996)

Stark JH; Smit JA; Lyons SF; Nel M; Burke J; Mokoena T
Interprimate stem cell transplantation.
Transplant Proc 28: 845-6 (1996)

Staszewski S; Loveday C; Picazo JJ; Dellarnonica P; Skinhoj P; Johnson MA; Danner SA; Harrigan PR; Hill AM; Verity L; McDade H
Safety and efficacy of lamivudine-zidovudine combination therapy in zidovudine-experienced patients. A randomized controlled comparison with zidovudine monotherapy. Lamivudine European HIV Working Group.
JAMA 276: 111-7 (1996)

Staszewski S; Miller V; Rehmet S; Stark T; De Cree J; De Brabander M; Peeters M; Andries K; Moeremans M; De Raeymaeker M; Pearce G; Van den Broeck R; Verbiest W; Stoffels P
Virological and immunological analysis of a triple combination pilot study with loviride, lamivudine and zidovudine in HIV-1-infected patients.
AIDS 10: F1-7 (1996)

Staszewski S; Miller V; Rehmet S; Stark T; De Cree J; De Brabander M; Peeters M; Andries K; Moeremans M; De Raeymaeker M; Pearce G; Van den Broeck R; Verbiest W; Stoffels P
Virological and immunological analysis of a triple combination pilot study with loviride, lamivudine and zidovudine in HIV-1-infected patients.
AIDS 10: F1-7 (1996)

Stephenson J
New anti-HIV drugs and treatment strategies buoy AIDS researchers [news]
JAMA 275: 579-80 (1996)

Stevens SW; Griffith JD
Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration.
J Virol 70: 6459-62 (1996)

Stoeckle MY; Douglas RG Jr
Infectious diseases.
JAMA 275: 1816-7 (1996)

Strack PR; Frey MW; Rizzo CJ; Cordova B; George HJ; Meade R; Ho SP; Corman J; Tritch R; Korant BD
Apoptosis mediated by HIV protease is preceded by cleavage of Bcl-2.
Proc Natl Acad Sci U S A 93: 9571-6 (1996)

Strack PR; Frey MW; Rizzo CJ; Cordova B; George HJ; Meade R; Ho SP; Corman J; Tritch R; Korant BD
Apoptosis mediated by HIV protease is preceded by cleavage of Bcl-2.
Proc Natl Acad Sci U S A 93: 9571-6 (1996)

Strathdee SA; Hogg RS; O'Shaughnessy MV; Montaner JS; Schechter MT
A decade of research on the natural history of HIV infection: Part 2. Cofactors.
Clin Invest Med 19: 121-30 (1996)

Sullivan DJ; Henman MC; Moran GP; O'Neill LC; Bennett DE; Shanley DB; Coleman DC
Molecular genetic approaches to identification, epidemiology and taxonomy of non-albicans Candida species.
J Med Microbiol 44: 399-408 (1996)

Susal C; Kirschfink M; Kropelin M; Daniel V; Opelz G
Identification of complement activation sites in human immunodeficiency virus type-1 glycoprotein gp120.
Blood 87: 2329-36 (1996)

Sutthent R; Foongladda S; Likanonskul S; Tunsupasawasdeekul S; Roongpisuthipong A; Chearskul S; Raktham S; Louisirirotchanakul S; Wasi C
Detection of HIV-1 proviral DNA by polymerase chain reaction: a preliminary study in Bangkok.
J Med Assoc Thai 79: 142-8 (1996)

Suzue K; Young RA
Adjuvant-free hsp70 fusion protein system elicits humoral and cellular immune responses to HIV-1 p24.
J Immunol 156: 873-9 (1996)

Swindells S; Baldwin T; Kelly C; Baca-Regen L; Loomis L; Post D; Brichacek B; Stevenson M; Dominguez EA; Reddy R; Klein R; Liao MJ; Testa D; McDonald T; Bellanti J; Skurkovich S; Gendelman HE
Regulation and characterization of the interferon-alpha present in patients with advanced human immunodeficiency virus type 1 disease.
J Interferon Cytokine Res 16: 127-37 (1996)

Swindells S; Baldwin T; Kelly C; Baca-Regen L; Loomis L; Post D; Brichacek B; Stevenson M; Dominguez EA; Reddy R; Klein R; Liao MJ; Testa D; McDonald T; Bellanti J; Skurkovich S; Gendelman HE
Regulation and characterization of the interferon-alpha present in patients with advanced human immunodeficiency virus type 1 disease.
J Interferon Cytokine Res 16: 127-37 (1996)

Symensma TL; Giver L; Zapp M; Takle GB; Ellington AD
RNA aptamers selected to bind human immunodeficiency virus type 1 Rev in vitro are Rev responsive in vivo.
J Virol 70: 179-87 (1996)

Szeltner Z; Polgar L
Conformational stability and catalytic activity of HIV-1 protease are both enhanced at high salt concentration.
J Biol Chem 271: 5458-63 (1996)

Tabernero C; Zolotukhin AS; Valentin A; Pavlakis GN; Felber BK
The posttranscriptional control element of the simian retrovirus type 1 forms an extensive RNA secondary structure necessary for its function.
J Virol 70: 5998-6011 (1996)

Takahashi H; Nakagawa Y; Leggatt GR; Ishida Y; Saito T; Yokomuro K; Berzofsky JA
Inactivation of human immunodeficiency virus (HIV)-1 envelope-specific CD8+ cytotoxic T lymphocytes by free antigenic peptide: a self-veto mechanism?
J Exp Med 183: 879-89 (1996)

Takahashi K; Wesselingh SL; Griffin DE; McArthur JC; Johnson RT; Glass JD
Localization of HIV-1 in human brain using polymerase chain reaction/in situ hybridization and immunocytochemistry.
Ann Neurol 39: 705-11 (1996)

Takatsuki K; Obaru K; Yoshimura K; Matsushita S
[Gene therapy for AIDS: current trends]
Nippon Rinsho 54: 233-41 (1996)

Tamamura H; Otaka A; Murakami T; Ishihara T; Ibuka T; Waki M; Matsumoto A; Yamamoto N; Fujii N
Interaction of an anti-HIV peptide, T22, with gp120 and CD4.
Biochem Biophys Res Commun 219: 555-9 (1996)

Tamma SL; Sundaram SK; Lev M; Coico RF
Inhibition of sphingolipid synthesis down-modulates CD4 expression by peripheral blood T lymphocytes and T lymphoma cells.
Biochem Biophys Res Commun 220: 916-21 (1996)

Tange TO; Jensen TH; Kjems J
In vitro interaction between human immunodeficiency virus type 1 Rev protein and splicing factor ASF/SF2-associated protein, p32.
J Biol Chem 271: 10066-72 (1996)

Tao J; Frankel AD
Arginine-binding RNAs resembling TAR identified by in vitro selection.
Biochemistry 35: 2229-38 (1996)

Taveira NC; Ferreira MO; Pereira JM
Amplification of full-length HIV-2 envelope genes.
Mol Cell Probes 10: 91-8 (1996)

Thaisrivongs S; Janakiraman MN; Chong KT; Tomich PK; Dolak LA; Turner SR; Strohbach JW; Lynn JC; Horng MM; Hinshaw RR; Watenpaugh KD
Structure-based design of novel HIV protease inhibitors: sulfonamide-containing 4-hydroxycoumarins and 4-hydroxy-2-pyrones as potent non-peptidic inhibitors.
J Med Chem 39: 2400-10 (1996)

Thiel G; Petersohn D; Schoch S
pHIVTATA-CAT, a versatile vector to study transcriptional regulatory elements in mammalian cells.
Gene 168: 173-6 (1996)

Thiesen HJ
From repression domains to designer zinc finger proteins: a novel strategy of intracellular immunization against HIV.
Gene Expr 5: 229-43 (1996)

Thiesen HJ
From repression domains to designer zinc finger proteins: a novel strategy of intracellular immunization against HIV.
Gene Expr 5: 229-43 (1996)

Thompson J; Pope T; Tung JS; Chan C; Hollis G; Mark G; Johnson KS
Affinity maturation of a high-affinity human monoclonal antibody against the third hypervariable loop of human immunodeficiency virus: use of phage display to improve affinity and broaden strain reactivity.
J Mol Biol 256: 77-88 (1996)

Thornton AM; Buller RM; DeVico AL; Wang IM; Ozato K
Inhibition of human immunodeficiency virus type 1 and vaccinia virus infection by a dominant negative factor of the interferon regulatory factor family expressed in monocytic cells.
Proc Natl Acad Sci U S A 93: 383-7 (1996)

Thrall SH; Reinstein J; Wohrl BM; Goody RS
Evaluation of human immunodeficiency virus type 1 reverse transcriptase primer tRNA binding by fluorescence spectroscopy: specificity and comparison to primer/template binding.
Biochemistry 35: 4609-18 (1996)

Toggas SM; Mucke L
Transgenic models in the study of AIDS dementia complex.
Curr Top Microbiol Immunol 206: 223-41 (1996)

Torres BA; Tanabe T; Johnson HM
Characterization of Nef-induced CD4 T cell proliferation.
Biochem Biophys Res Commun 225: 54-61 (1996)

Torres BA; Tanabe T; Yamamoto JK; Johnson HM
HIV encodes for its own CD4 T-cell superantigen mitogen.
Biochem Biophys Res Commun 225: 672-8 (1996)

Torres BA; Tanabe T; Yamamoto JK; Johnson HM
HIV encodes for its own CD4 T-cell superantigen mitogen.
Biochem Biophys Res Commun 225: 672-8 (1996)

Tozser J; Bagossi P; Weber IT; Copeland TD; Oroszlan S
Comparative studies on the substrate specificity of avian myeloblastosis virus proteinase and lentiviral proteinases.
J Biol Chem 271: 6781-8 (1996)

Trkola A; Purtscher M; Muster T; Ballaun C; Buchacher A; Sullivan N; Srinivasan K; Sodroski J; Moore JP; Katinger H
Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1.
J Virol 70: 1100-8 (1996)

Trujillo JR; Wang WK; Lee TH; Essex M
Identification of the envelope V3 loop as a determinant of a CD4-negative neuronal cell tropism for HIV-1.
Virology 217: 613-7 (1996)

Tsang SX; Morris GF; Lu M; Morris CB
TATA-dependent repression of human immunodeficiency virus type-1 transcription by the adenovirus E1A 243R oncoprotein.
Oncogene 12: 819-26 (1996)

Tsolaki AG; Miller RF; Underwood AP; Banerji S; Wakefield AE
Genetic diversity at the internal transcribed spacer regions of the rRNA operon among isolates of Pneumocystis carinii from AIDS patients with recurrent pneumonia.
J Infect Dis 174: 141-56 (1996)

Tsotinis A; Calogeropoulou T; Koufaki M; Souli C; Balzarini J; De Clercq E; Makriyannis A
Synthesis and antiretroviral evaluation of new alkoxy and aryloxy phosphate derivatives of 3'-azido-3'-deoxythymidine.
J Med Chem 39: 3418-22 (1996)

Tsuchiya H; Iseda T; Hino O
Identification of a novel protein (VBP-1) binding to the von Hippel-Lindau (VHL) tumor suppressor gene product.
Cancer Res 56: 2881-5 (1996)

Tumminia SJ; Jonak GJ; Focht RJ; Cheng YS; Russell P
Cataractogenesis in transgenic mice containing the HIV-1 protease linked to the lens alpha A-crystallin promoter.
J Biol Chem 271: 425-31 (1996)

Tummino PJ; Scholten JD; Harvey PJ; Holler TP; Maloney L; Gogliotti R; Domagala J; Hupe D
The in vitro ejection of zinc from human immunodeficiency virus (HIV) type 1 nucleocapsid protein by disulfide benzamides with cellular anti-HIV activity.
Proc Natl Acad Sci U S A 93: 969-73 (1996)

Tung FY; Tung MH
Characterization of antisense RNA-mediated inhibition of SIV replication.
J Med Virol 48: 321-5 (1996)

Tung FY; Tung MH
Characterization of antisense RNA-mediated inhibition of SIV replication.
J Med Virol 48: 321-5 (1996)

Turpin JA; Terpening SJ; Schaeffer CA; Yu G; Glover CJ; Felsted RL; Sausville EA; Rice WG
Inhibitors of human immunodeficiency virus type 1 zinc fingers prevent normal processing of gag precursors and result in the release of noninfectious virus particles.
J Virol 70: 6180-9 (1996)

Tyagi S; Landegren U; Tazi M; Lizardi PM; Kramer FR
Extremely sensitive, background-free gene detection using binary probes and beta replicase.
Proc Natl Acad Sci U S A 93: 5395-400 (1996)

Uchiumi F; Maruta H; Inoue J; Yamamoto T; Tanuma S
Inhibitory effect of tannic acid on human immunodeficiency virus promoter activity induced by 12-O-tetra decanoylphorbol-13-acetate in Jurkat T-cells.
Biochem Biophys Res Commun 220: 411-7 (1996)

Ueland PM; Mansoor MA; Guttormsen AB; Muller F; Aukrust P; Refsum H; Svardal AM
Reduced, oxidized and protein-bound forms of homocysteine and other aminothiols in plasma comprise the redox thiol status–a possible element of the extracellular antioxidant defense system.
J Nutr 126: 1281S-4S (1996)

Uhlikova T; Konvalinka J; Pichova I; Soucek M; Krausslich HG; Vondrasek J
A modular approach to HIV-1 proteinase inhibitor design.
Biochem Biophys Res Commun 222: 38-43 (1996)

Ulich C; Harrich D; Estes P; Gaynor RB
Inhibition of human immunodeficiency virus type 1 replication is enhanced by a combination of transdominant Tat and Rev proteins.
J Virol 70: 4871-6 (1996)

Urban J; Qabar M; Sia C; Klein M; Kahn M
Sculpted immunogens; B-cell epitope optimization using constrained secondary structure libraries.
Bioorg Med Chem 4: 673-6 (1996)

Urban J; Qabar M; Sia C; Klein M; Kahn M
Sculpted immunogens; B-cell epitope optimization using constrained secondary structure libraries.
Bioorg Med Chem 4: 673-6 (1996)

Urnovitz HB; Murphy WH
Human endogenous retroviruses: nature, occurrence, and clinical implications in human disease.
Clin Microbiol Rev 9: 72-99 (1996)

Vago L; Cinque P; Sala E; Nebuloni M; Caldarelli R; Racca S; Ferrante P; Trabottoni G; Costanzi G
JCV-DNA and BKV-DNA in the CNS tissue and CSF of AIDS patients and normal subjects. Study of 41 cases and review of the literature.
J Acquir Immune Defic Syndr Hum Retrovirol 12: 139-46 (1996)

Valerie K; Laster WS; Cheng L; Kirkham JC; Reavey P; Kuemmere NB
Signal transduction and HIV transcriptional activation after exposure to ultraviolet light and other DNA-damaging agents.
Photochem Photobiol 64: 280-5 (1996)

Valerie K; Laster WS; Cheng L; Kirkham JC; Reavey P; Kuemmere NB
Signal transduction and HIV transcriptional activation after exposure to ultraviolet light and other DNA-damaging agents.
Photochem Photobiol 64: 280-5 (1996)

Valvatne H; Szilvay AM; Helland DE
A monoclonal antibody defines a novel HIV type 1 Tat domain involved in trans-cellular trans-activation.
AIDS Res Hum Retroviruses 12: 611-9 (1996)

Valvatne H; Szilvay AM; Helland DE
A monoclonal antibody defines a novel HIV type 1 Tat domain involved in trans-cellular trans-activation.
AIDS Res Hum Retroviruses 12: 611-9 (1996)

Van Lint C; Emiliani S; Ott M; Verdin E
Transcriptional activation and chromatin remodeling of the HIV-1 promoter in response to histone acetylation.
EMBO J 15: 1112-20 (1996)

Van Lint C; Emiliani S; Verdin E
The expression of a small fraction of cellular genes is changed in response to histone hyperacetylation.
Gene Expr 5: 245-53 (1996)

Van Lint C; Emiliani S; Verdin E
The expression of a small fraction of cellular genes is changed in response to histone hyperacetylation.
Gene Expr 5: 245-53 (1996)

Vandamme AM; Liu HF; Van Brussel M; De Meurichy W; Desmyter J; Goubau P
The presence of a divergent T-lymphotropic virus in a wild-caught pygmy chimpanzee (Pan paniscus) supports an African origin for the human T-lymphotropic/simian T-lymphotropic group of viruses.
J Gen Virol 77 (Pt 5): 1089-99 (1996)

Vandamme AM; Liu HF; Van Brussel M; De Meurichy W; Desmyter J; Goubau P
The presence of a divergent T-lymphotropic virus in a wild-caught pygmy chimpanzee (Pan paniscus) supports an African origin for the human T-lymphotropic/simian T-lymphotropic group of viruses.
J Gen Virol 77 (Pt 5): 1089-99 (1996)

Vanden Haesevelde MM; Peeters M; Jannes G; Janssens W; van der Groen G; Sharp PM; Saman E
Sequence analysis of a highly divergent HIV-1-related lentivirus isolated from a wild captured chimpanzee.
Virology 221: 346-50 (1996)

Vanden Haesevelde MM; Peeters M; Jannes G; Janssens W; van der Groen G; Sharp PM; Saman E
Sequence analysis of a highly divergent HIV-1-related lentivirus isolated from a wild captured chimpanzee.
Virology 221: 346-50 (1996)

Vella S; Galluzzo C; Giannini G; Pirillo MF; Duncan I; Jacobsen H; Andreoni M; Sarmati L; Ercoli L
Saquinavir/zidovudine combination in patients with advanced HIV infection and no prior antiretroviral therapy: CD4+ lymphocyte/plasma RNA changes, and emergence of HIV strains with reduced phenotypic sensitivity.
Antiviral Res 29: 91-3 (1996)

Vella S; Galluzzo C; Giannini G; Pirillo MF; Duncan I; Jacobsen H; Andreoni M; Sarmati L; Ercoli L
Saquinavir/zidovudine combination in patients with advanced HIV infection and no prior antiretroviral therapy: CD4+ lymphocyte/plasma RNA changes, and emergence of HIV strains with reduced phenotypic sensitivity.
Antiviral Res 29: 91-3 (1996)

Verkhivker GM; Rejto PA
A mean field model of ligand-protein interactions: implications for the structural assessment of human immunodeficiency virus type 1 protease complexes and receptor-specific binding.
Proc Natl Acad Sci U S A 93: 60-4 (1996)

Vessey SJ; Bell JI; Jakobsen BK
A functionally significant allelic polymorphism in a T cell receptor V beta gene segment.
Eur J Immunol 26: 1660-3 (1996)

Viswanadhan VN; Reddy MR; Wlodawer A; Varney MD; Weinstein JN
An approach to rapid estimation of relative binding affinities of enzyme inhibitors: application to peptidomimetic inhibitors of the human immunodeficiency virus type 1 protease.
J Med Chem 39: 705-12 (1996)

Voelker R
Gene therapy for HIV [news]
JAMA 275: 1533 (1996)

Voevodin A; Crandall KA; Seth P; al Mufti S
HIV type 1 subtypes B and C from new regions of India and Indian and Ethiopian expatriates in Kuwait.
AIDS Res Hum Retroviruses 12: 641-3 (1996)

Voevodin A; Crandall KA; Seth P; al Mufti S
HIV type 1 subtypes B and C from new regions of India and Indian and Ethiopian expatriates in Kuwait.
AIDS Res Hum Retroviruses 12: 641-3 (1996)

Volberding PA
HIV quantification: clinical applications.
Lancet 347: 71-3 (1996)

Vollenweider F; Benjannet S; Decroly E; Savaria D; Lazure C; Thomas G; Chretien M; Seidah NG
Comparative cellular processing of the human immunodeficiency virus (HIV-1) envelope glycoprotein gp160 by the mammalian subtilisin/kexin-like convertases.
Biochem J 314 (Pt 2): 521-32 (1996)

Volsky DJ; Simm M; Shahabuddin M; Li G; Chao W; Potash MJ
Interference to human immunodeficiency virus type 1 infection in the absence of downmodulation of the principal virus receptor, CD4.
J Virol 70: 3823-33 (1996)

Volter C; He Y; Delius H; Roy-Burman A; Greenspan JS; Greenspan D; de Villiers EM
Novel HPV types present in oral papillomatous lesions from patients with HIV infection.
Int J Cancer 66: 453-6 (1996)

Vranken WF; Budesinsky M; Martins JC; Fant F; Boulez K; Gras-Masse H; Borremans FA
Conformational features of a synthetic cyclic peptide corresponding to the complete V3 loop of the RF HIV-1 strain in water and water/trifluoroethanol solutions.
Eur J Biochem 236: 100-8 (1996)

Vu HM; de Lorimier R; Moody MA; Haynes BF; Spicer LD
Conformational preferences of a chimeric peptide HIV-1 immunogen from the C4-V3 domains of gp120 envelope protein of HIV-1 CAN0A based on solution NMR: comparison to a related immunogenic peptide from HIV-1 RF.
Biochemistry 35: 5158-65 (1996)

Vzorov AN; Compans RW
Assembly and release of SIV env proteins with full-length or truncated cytoplasmic domains.
Virology 221: 22-33 (1996)

Wada N; Ohara N; Kameoka M; Nishino Y; Matsumoto S; Nishiyama T; Naito M; Yukitake H; Okada Y; Ikuta K; Yamada T
Long-lasting immune response induced by recombinant bacillus Calmette-Guerin (BCG) secretion system.
Scand J Immunol 43: 202-9 (1996)

Wagener S; Norley S; zur Megede J; Kurth R; Cichutek K
Induction of antibodies against SIV antigens after intramuscular nucleic acid inoculation using complex expression constructs.
J Biotechnol 44: 59-65 (1996)

Wagener S; Norley S; zur Megede J; Kurth R; Cichutek K
Induction of antibodies against SIV antigens after intramuscular nucleic acid inoculation using complex expression constructs.
J Biotechnol 44: 59-65 (1996)

Wagner R; Deml L; Schirmbeck R; Niedrig M; Reimann J; Wolf H
Construction, expression, and immunogenicity of chimeric HIV-1 virus-like particles.
Virology 220: 128-40 (1996)

Wain-Hobson S
Running the gamut of retroviral variation.
Trends Microbiol 4: 135-41 (1996)

Wain-Hobson S
Running the gamut of retroviral variation.
Trends Microbiol 4: 135-41 (1996)

Wain-Hobson S
Running the gamut of retroviral variation.
Trends Microbiol 4: 135-41 (1996)

Wainberg MA; Drosopoulos WC; Salomon H; Hsu M; Borkow G; Parniak M; Gu Z; Song Q; Manne J; Islam S; Castriota G; Prasad VR
Enhanced fidelity of 3TC-selected mutant HIV-1 reverse transcriptase.
Science 271: 1282-5 (1996)

Wakefield JK; Kang SM; Morrow CD
Construction of a type 1 human immunodeficiency virus that maintains a primer binding site complementary to tRNA(His).
J Virol 70: 966-75 (1996)

Wakefield JK; Morrow CD
Mutations within the primer binding site of the human immunodeficiency virus type 1 define sequence requirements essential for reverse transcription.
Virology 220: 290-8 (1996)

Wakrim L; Le Grand R; Vaslin B; Cheret A; Matheux F; Theodoro F; Roques P; Nicol-Jourdain I; Dormont D
Superinfection of HIV-2-preinfected macaques after rectal exposure to a primary isolate of SIVmac251.
Virology 221: 260-70 (1996)

Wakrim L; Le Grand R; Vaslin B; Cheret A; Matheux F; Theodoro F; Roques P; Nicol-Jourdain I; Dormont D
Superinfection of HIV-2-preinfected macaques after rectal exposure to a primary isolate of SIVmac251.
Virology 221: 260-70 (1996)

Wallace RJ Jr; Brown BA; Griffith DE; Girard WM; Murphy DT
Clarithromycin regimens for pulmonary Mycobacterium avium complex. The first 50 patients [see comments]
Am J Respir Crit Care Med 153: 1766-72 (1996)

Wan M; Takagi M; Loh BN; Xu XZ; Imanaka T
Autoprocessing: an essential step for the activation of HIV-1 protease.
Biochem J 316 (Pt 2): 569-73 (1996)

Wang H; Ben-Naim A
A possible involvement of solvent-induced interactions in drug design.
J Med Chem 39: 1531-9 (1996)

Wang HK; Bastow KF; Cosentino LM; Lee KH
Antitumor agents. 166. Synthesis and biological evaluation of 5,6,7,8-substituted-2-phenylthiochromen-4-ones.
J Med Chem 39: 1975-80 (1996)

Wang J; Cai X; Rivas G; Shiraishi H; Farias PA; Dontha N
DNA electrochemical biosensor for the detection of short DNA sequences related to the human immunodeficiency virus.
Anal Chem 68: 2629-34 (1996)

Wang S; Milne GW; Yan X; Posey IJ; Nicklaus MC; Graham L; Rice WG
Discovery of novel, non-peptide HIV-1 protease inhibitors by pharmacophore searching.
J Med Chem 39: 2047-54 (1996)

Wang WK; Essex M; Lee TH
Single amino acid substitution in constant region 1 or 4 of gp120 causes the phenotype of a human immunodeficiency virus type 1 variant with mutations in hypervariable regions 1 and 2 to revert.
J Virol 70: 607-11 (1996)

Wang WK; Essex M; McLane MF; Mayer KH; Hsieh CC; Brumblay HG; Seage G; Lee TH
Pattern of gp120 sequence divergence linked to a lack of clinical progression in human immunodeficiency virus type 1 infection.
Proc Natl Acad Sci U S A 93: 6693-7 (1996)

Wang YX; Freedberg DI; Yamazaki T; Wingfield PT; Stahl SJ; Kaufman JD; Kiso Y; Torchia DA
Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272.
Biochemistry 35: 9945-50 (1996)

Wang Z; Morris GF; Rice AP; Xiong W; Morris CB
Wild-type and transactivation-defective mutants of human immunodeficiency virus type 1 Tat protein bind human TATA-binding protein in vitro.
J Acquir Immune Defic Syndr Hum Retrovirol 12: 128-38 (1996)

Wang Z; Rana TM
RNA conformation in the Tat-TAR complex determined by site-specific photo-cross-linking.
Biochemistry 35: 6491-9 (1996)

Wang Z; Wang X; Rana TM
Protein orientation in the Tat-TAR complex determined by psoralen photocross-linking.
J Biol Chem 271: 16995-8 (1996)

Watkins BA; Davis AE; Fiorentini S; di Marzo Veronese F; Reitz MS Jr
Evidence for distinct contributions of heavy and light chains to restriction of antibody recognition of the HIV-1 principal neutralization determinant.
J Immunol 156: 1676-83 (1996)

Weiss RA
HIV receptors and the pathogenesis of AIDS.
Science 272: 1885-6 (1996)

Weissenhorn W; Wharton SA; Calder LJ; Earl PL; Moss B; Aliprandis E; Skehel JJ; Wiley DC
The ectodomain of HIV-1 env subunit gp41 forms a soluble, alpha-helical, rod-like oligomer in the absence of gp120 and the N-terminal fusion peptide.
EMBO J 15: 1507-14 (1996)

Welker R; Kottler H; Kalbitzer HR; Krausslich HG
Human immunodeficiency virus type 1 Nef protein is incorporated into virus particles and specifically cleaved by the viral proteinase.
Virology 219: 228-36 (1996)

Wieland U; Suhr H; Salzberger B; Eggers HJ; Braun RW; Kuhn JE
Quantification of HIV-1 proviral DNA and analysis of genomic diversity by polymerase chain reaction and temperature gradient gel electrophoresis.
J Virol Methods 57: 127-39 (1996)

Wieland U; Suhr H; Salzberger B; Eggers HJ; Braun RW; Kuhn JE
Quantification of HIV-1 proviral DNA and analysis of genomic diversity by polymerase chain reaction and temperature gradient gel electrophoresis.
J Virol Methods 57: 127-39 (1996)

Wilk T; Pfeiffer T; Bukovsky A; Moldenhauer G; Bosch V
Glycoprotein incorporation and HIV-1 infectivity despite exchange of the gp160 membrane-spanning domain.
Virology 218: 269-74 (1996)

Willard-Gallo KE; Delmelle-Wibaut C; Segura-Zapata I; Janssens M; Willems L; Kettmann R
Modulation of CD3-gamma gene expression after HIV type 1 infection of the WE17/10 T cell line is progressive and occurs in concert with decreased production of viral p24 antigen.
AIDS Res Hum Retroviruses 12: 715-25 (1996)

Willard-Gallo KE; Delmelle-Wibaut C; Segura-Zapata I; Janssens M; Willems L; Kettmann R
Modulation of CD3-gamma gene expression after HIV type 1 infection of the WE17/10 T cell line is progressive and occurs in concert with decreased production of viral p24 antigen.
AIDS Res Hum Retroviruses 12: 715-25 (1996)

Williams RD; Lee BA; Jackson SP; Proudfoot NJ
Activation domains of transcription factors mediate replication dependent transcription from a minimal HIV-1 promoter.
Nucleic Acids Res 24: 549-57 (1996)

Williams SJ; Schwer C; Krishnarao AS; Heid C; Karger BL; Williams PM
Quantitative competitive polymerase chain reaction: analysis of amplified products of the HIV-1 gag gene by capillary electrophoresis with laser-induced fluorescence detection.
Anal Biochem 236: 146-52 (1996)

Wilson JE; Aulabaugh A; Caligan B; McPherson S; Wakefield JK; Jablonski S; Morrow CD; Reardon JE; Furman PA
Human immunodeficiency virus type-1 reverse transcriptase. Contribution of Met-184 to binding of nucleoside 5'-triphosphate.
J Biol Chem 271: 13656-62 (1996)

Winter H; Maeda Y; Mitsuya H; Zemlicka J
Phosphodiester amidates of allenic nucleoside analogues: anti-HIV activity and possible mechanism of action.
J Med Chem 39: 3300-6 (1996)

Wlassoff WA; Dymshits GM; Lavrik OI
A model for DNA polymerase translocation: worm-like movement of DNA within the binding cleft.
FEBS Lett 390: 6-9 (1996)

Woffendin C; Ranga U; Yang Z; Xu L; Nabel GJ
Expression of a protective gene-prolongs survival of T cells in human immunodeficiency virus-infected patients.
Proc Natl Acad Sci U S A 93: 2889-94 (1996)

Wolfe AL; Felock PJ; Hastings JC; Blau CU; Hazuda DJ
The role of manganese in promoting multimerization and assembly of human immunodeficiency virus type 1 integrase as a catalytically active complex on immobilized long terminal repeat substrates.
J Virol 70: 1424-32 (1996)

Wolinsky SM; Korber BT; Neumann AU; Daniels M; Kunstman KJ; Whetsell AJ; Furtado MR; Cao Y; Ho DD; Safrit JT; Koup RA
Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection [see comments]
Science 272: 537-42 (1996)

Wolthers KC; Otto SA; Lens SM; Kolbach DN; van Lier RA; Miedema F; Meyaard L
Increased expression of CD80, CD86 and CD70 on T cells from HIV-infected individuals upon activation in vitro: regulation by CD4+ T cells.
Eur J Immunol 26: 1700-6 (1996)

Wolthers KC; Otto SA; Lens SM; Kolbach DN; van Lier RA; Miedema F; Meyaard L
Increased expression of CD80, CD86 and CD70 on T cells from HIV-infected individuals upon activation in vitro: regulation by CD4+ T cells.
Eur J Immunol 26: 1700-6 (1996)

Wondrak EM; Nashed NT; Haber MT; Jerina DM; Louis JM
A transient precursor of the HIV-1 protease. Isolation, characterization, and kinetics of maturation.
J Biol Chem 271: 4477-81 (1996)

Wong MT; Dolan MJ; Kozlow E; Doe R; Melcher GP; Burke DS; Boswell RN; Vahey M
Patterns of virus burden and T cell phenotype are established early and are correlated with the rate of disease progression in human immunodeficiency virus type 1-infected persons.
J Infect Dis 173: 877-87 (1996)

Wu W; Palaniappan C; Bambara RA; Fay PJ
Differences in mutagenesis during minus strand, plus strand and strand transfer (recombination) synthesis of the HIV-1 gene in vitro.
Nucleic Acids Res 24: 1710-8 (1996)

Wu X; Liu H; Xiao H; Conway JA; Kappes JC
Inhibition of human and simian immunodeficiency virus protease function by targeting Vpx-protease-mutant fusion protein into viral particles.
J Virol 70: 3378-84 (1996)

Wu X; Liu H; Xiao H; Conway JA; Kappes JC
Inhibition of human and simian immunodeficiency virus protease function by targeting Vpx-protease-mutant fusion protein into viral particles.
J Virol 70: 3378-84 (1996)

Wu X; Liu H; Xiao H; Kappes JC
Proteolytic activity of human immunodeficiency virus Vpr- and Vpx-protease fusion proteins.
Virology 219: 307-13 (1996)

Wu Y; Duan L; Zhu M; Hu B; Kubota S; Bagasra O; Pomerantz RJ
Binding of intracellular anti-Rev single chain variable fragments to different epitopes of human immunodeficiency virus type 1 rev: variations in viral inhibition.
J Virol 70: 3290-7 (1996)

Wu-Baer F; Lane WS; Gaynor RB
Identification of a group of cellular cofactors that stimulate the binding of RNA polymerase II and TRP-185 to human immunodeficiency virus 1 TAR RNA.
J Biol Chem 271: 4201-8 (1996)

Wyatt JR; Davis PW; Freier SM
Kinetics of G-quartet-mediated tetramer formation.
Biochemistry 35: 8002-8 (1996)

Wyss-Coray T; Masliah E; Toggas SM; Rockenstein EM; Brooker MJ; Lee HS; Mucke L
Dysregulation of signal transduction pathways as a potential mechanism of nervous system alterations in HIV-1 gp120 transgenic mice and humans with HIV-1 encephalitis.
J Clin Invest 97: 789-98 (1996)

Xu M; Sharma V; Bryans M; Wilkie NM
Identification of a new member of the MNP transcription factor family in differentiated HL60 cells.
Biochem Biophys Res Commun 226: 488-94 (1996)

Xu M; Sharma V; Bryans M; Wilkie NM
Identification of a new member of the MNP transcription factor family in differentiated HL60 cells.
Biochem Biophys Res Commun 226: 488-94 (1996)

Yamada O; Kraus G; Luznik L; Yu M; Wong-Staal F
A chimeric human immunodeficiency virus type 1 (HIV-1) minimal Rev response element-ribozyme molecule exhibits dual antiviral function and inhibits cell-cell transmission of HIV-1.
J Virol 70: 1596-601 (1996)

Yamada O; Kraus G; Sargueil B; Yu Q; Burke JM; Wong-Staal F
Conservation of a hairpin ribozyme sequence in HIV-1 is required for efficient viral replication.
Virology 220: 361-6 (1996)

Yan ZX; Meyer TF
Mixed population approach for vaccination with live recombinant Salmonella strains.
J Biotechnol 44: 197-201 (1996)

Yan ZX; Meyer TF
Mixed population approach for vaccination with live recombinant Salmonella strains.
J Biotechnol 44: 197-201 (1996)

Yang C; Compans RW
Analysis of the cell fusion activities of chimeric simian immunodeficiency virus-murine leukemia virus envelope proteins: inhibitory effects of the R peptide.
J Virol 70: 248-54 (1996)

Yang G; Song Q; Charles M; Drosopoulos WC; Arnold E; Prasad VR
Use of chimeric human immunodeficiency virus types 1 and 2 reverse transcriptases for structure-function analysis and for mapping susceptibility to nonnucleoside inhibitors.
J Acquir Immune Defic Syndr Hum Retrovirol 11: 326-33 (1996)

Yang OO; Kalams SA; Rosenzweig M; Trocha A; Jones N; Koziel M; Walker BD; Johnson RP
Efficient lysis of human immunodeficiency virus type 1-infected cells by cytotoxic T lymphocytes.
J Virol 70: 5799-806 (1996)

Yang X; Goncalves J; Gabuzda D
Phosphorylation of Vif and its role in HIV-1 replication.
J Biol Chem 271: 10121-9 (1996)

Yankulov KY; Pandes M; McCracken S; Bouchard D; Bentley DL
TFIIH functions in regulating transcriptional elongation by RNA polymerase II in Xenopus oocytes.
Mol Cell Biol 16: 3291-9 (1996)

Yano K; Toki S; Nakanishi S; Ochiai K; Ando K; Yoshida M; Matsuda Y; Yamasaki M
MS-271, a novel inhibitor of calmodulin-activated myosin light chain kinase from Streptomyces sp.–I. Isolation, structural determination and biological properties of MS-271.
Bioorg Med Chem 4: 115-20 (1996)

Yoo J; Chen H; Kraus T; Hirsch D; Polyak S; George I; Sperber K
Altered cytokine production and accessory cell function after HIV-1 infection.
J Immunol 157: 1313-20 (1996)

Yoshimura K; Matsushita S; Hayashi A; Takatsuki K
Relationship of HIV-1 envelope V2 and V3 sequences of the primary isolates to the viral phenotype.
Microbiol Immunol 40: 277-87 (1996)

Yu H; Rabson AB; Kaul M; Ron Y; Dougherty JP
Inducible human immunodeficiency virus type 1 packaging cell lines.
J Virol 70: 4530-7 (1996)

Yusupova G; Lanchy JM; Yusupov M; Keith G; Le Grice SF; Ehresmann C; Ehresmann B; Marquet R
Primer selection by HIV-1 reverse transcriptase on RNA-tRNA(3Lys) and DNA-tRNA(3Lys) hybrids.
J Mol Biol 261: 315-21 (1996)

Yusupova G; Lanchy JM; Yusupov M; Keith G; Le Grice SF; Ehresmann C; Ehresmann B; Marquet R
Primer selection by HIV-1 reverse transcriptase on RNA-tRNA(3Lys) and DNA-tRNA(3Lys) hybrids.
J Mol Biol 261: 315-21 (1996)

Zakharova OD; Suturina OA; Gudima SO; Pokholok DK; Iamkovoi VI; Kochetkov SN; Nevinskii GA
[Interaction of primers with mutant forms of human immunodeficiency virus reverse transcriptase]
Mol Biol (Mosk) 30: 231-40 (1996)

Zakharova OD; Suturina OA; Gudima SO; Pokholok DK; Iamkovoi VI; Kochetkov SN; Nevinskii GA
[Interaction of primers with mutant forms of human immunodeficiency virus reverse transcriptase]
Mol Biol (Mosk) 30: 231-40 (1996)

Zauli G; Vitale M; Gibellini D; Capitani S
Inhibition of purified CD34+ hematopoietic progenitor cells by human immunodeficiency virus 1 or gp120 mediated by endogenous transforming growth factor beta 1.
J Exp Med 183: 99-108 (1996)

Zemmel RW; Kelley AC; Karn J; Butler PJ
Flexible regions of RNA structure facilitate co-operative Rev assembly on the Rev-response element.
J Mol Biol 258: 763-77 (1996)

Zhang H; Dornadula G; Pomerantz RJ
Endogenous reverse transcription of human immunodeficiency virus type 1 in physiological microenviroments: an important stage for viral infection of nondividing cells.
J Virol 70: 2809-24 (1996)

Zhang H; Dornadula G; Wu Y; Havlir D; Richman DD; Pomerantz RJ
Kinetic analysis of intravirion reverse transcription in the blood plasma of human immunodeficiency virus type 1-infected individuals: direct assessment of resistance to reverse transcriptase inhibitors in vivo.
J Virol 70: 628-34 (1996)

Zhang WH; Hockley DJ; Nermut MV; Morikawa Y; Jones IM
Gag-Gag interactions in the C-terminal domain of human immunodeficiency virus type 1 p24 capsid antigen are essential for Gag particle assembly.
J Gen Virol 77 (Pt 4): 743-51 (1996)

Zhao S; Ooi SL; Yang FC; Pardee AB
Three methods for identification of true positive cloned cDNA fragment in differential display.
Biotechniques 20: 400-4 (1996)

Zhao Y; Cao J; O'Gorman MR; Yu M; Yogev R
Effect of human immunodeficiency virus type 1 protein R (vpr) gene expression on basic cellular function of fission yeast Schizosaccharomyces pombe.
J Virol 70: 5821-6 (1996)

Zheng NN; McQueen PW; Hurren L; Evans LA; Law MG; Forde S; Barker S; Cooper DA; Delaney SF
Changes in biologic phenotype of human immunodeficiency virus during treatment of patients with didanosine.
J Infect Dis 173: 1092-6 (1996)

Zhong W; Wang H; Herndier B; Ganem D
Restricted expression of Kaposi sarcoma-associated herpesvirus (human herpesvirus 8) genes in Kaposi sarcoma.
Proc Natl Acad Sci U S A 93: 6641-6 (1996)

Zhu GW; Mukherjee S; Sahni M; Narayan O; Stephens EB
Prolonged infection in rhesus macaques with simian immunodeficiency virus (SIVmac239) results in animal-specific and rarely tissue-specific selection of nef variants.
Virology 220: 522-9 (1996)

Zhu QY; Scarborough A; Polsky B; Chou TC
Drug combinations and effect parameters of zidovudine, stavudine, and nevirapine in standardized drug-sensitive and resistant HIV type 1 strains.
AIDS Res Hum Retroviruses 12: 507-17 (1996)

Zhu T; Wang N; Carr A; Nam DS; Moor-Jankowski R; Cooper DA; Ho DD
Genetic characterization of human immunodeficiency virus type 1 in blood and genital secretions: evidence for viral compartmentalization and selection during sexual transmission.
J Virol 70: 3098-107 (1996)

Zola H
Analysis of receptors for cytokines and growth factors in human disease.
Dis Markers 12: 225-40 (1996)

Zola H
Analysis of receptors for cytokines and growth factors in human disease.
Dis Markers 12: 225-40 (1996)

Zou JX; Luciw PA
The requirement for Vif of SIVmac is cell-type dependent.
J Gen Virol 77 (Pt 3): 427-34 (1996)

Zou JX; Luciw PA
The requirement for Vif of SIVmac is cell-type dependent.
J Gen Virol 77 (Pt 3): 427-34 (1996)

al-Ghusein H; Ball H; Igloi GL; Gbewonyo A; Coates AR; Mascagni P; Roberts MM
Chemically synthesised human immunodeficiency virus P7 nucleocapsid protein can self-assemble into particles and binds to a specific site on the tRNA(Lys,3) primer.
Biochem Biophys Res Commun 224: 191-8 (1996)

de Clercq E
Non-nucleoside reverse transcriptase inhibitors (NNRTIs) for the treatment of human immunodeficiency virus type 1 (HIV-1) infections: strategies to overcome drug resistance development.
Med Res Rev 16: 125-57 (1996)

de Jong AL; Green DM; Trial JA; Birdsall HH
Focal effects of mononuclear leukocyte transendothelial migration: TNF-alpha production by migrating monocytes promotes subsequent migration of lymphocytes.
J Leukoc Biol 60: 129-36 (1996)

de Jong MD; Veenstra J; Stilianakis NI; Schuurman R; Lange JM; de Boer RJ; Boucher CA
Host-parasite dynamics and outgrowth of virus containing a single K70R amino acid change in reverse transcriptase are responsible for the loss of human immunodeficiency virus type 1 RNA load suppression by zidovudine.
Proc Natl Acad Sci U S A 93: 5501-6 (1996)

de Ronde A; de Rooij ER; Coutinho RA; Goudsmit J
[Zidovudine-resistant HIV strains in intravenous drug users and homosexual men in Amsterdam]
Ned Tijdschr Geneeskd 140: 932-4 (1996)

de Vreese K; Kofler-Mongold V; Leutgeb C; Weber V; Vermeire K; Schacht S; Anne J; de Clercq E; Datema R; Werner G
The molecular target of bicyclams, potent inhibitors of human immunodeficiency virus replication.
J Virol 70: 689-96 (1996)

di Marzo Veronese F; Arnott D; Barnaba V; Loftus DJ; Sakaguchi K; Thompson CB; Salemi S; Mastroianni C; Sette A; Shabanowitz J; Hunt DF; Appella E
Autoreactive cytotoxic T lymphocytes in human immunodeficiency virus type 1-infected subjects.
J Exp Med 183: 2509-16 (1996)

el Kharroubi A; Martin MA
cis-acting sequences located downstream of the human immunodeficiency virus type 1 promoter affect its chromatin structure and transcriptional activity.
Mol Cell Biol 16: 2958-66 (1996)

van Baalen CA; Klein MR; Huisman RC; Dings ME; Kerkhof Garde SR; Geretti AM; Gruters R; van Els CA; Miedema F; Osterhaus AD
Fine-specificity of cytotoxic T lymphocytes which recognize conserved epitopes of the Gag protein of human immunodeficiency virus type 1.
J Gen Virol 77 (Pt 8): 1659-65 (1996)

van Gent DC; Mizuuchi K; Gellert M
Similarities between initiation of V(D)J recombination and retroviral integration [see comments]
Science 271: 1592-4 (1996)

van der Burg SH; Visseren MJ; Brandt RM; Kast WM; Melief CJ
Immunogenicity of peptides bound to MHC class I molecules depends on the MHC-peptide complex stability.
J Immunol 156: 3308-14 (1996)

van der Hoek W; van Oosterhout JJ; Ngoma M
An outbreak of dysentery in Zambia [letter]
S Afr Med J 86: 93-4 (1996)