

Investigation of the genetic and molecular
basis of HIV-1 cellular tropism *in vivo*

Matthew David Marsden

Doctor of Philosophy

The University of Edinburgh

2003



Contents

DECLARATION.....	I
ACKNOWLEDGEMENTS	II
ABBREVIATION LIST	III
LIST OF FIGURES	VII
LIST OF TABLES	IX
ABSTRACT.....	X
CHAPTER 1: GENERAL INTRODUCTION.....	1
1.1 HISTORICAL PERSPECTIVE.....	2
1.1.1 CLASSIFICATION AND MORPHOLOGY OF RETROVIRUSES	4
1.1.2 ORIGINS AND GLOBAL DIVERSITY OF HIV.....	6
1.2 THE HIV-1 VIRION	8
1.2.1 GENOMIC ORGANISATION OF HIV-1	10
1.2.2 STRUCTURAL PROTEINS	12
<i>Gag</i>	12
<i>Pol</i>	14
<i>Env</i>	16
1.2.3 REGULATORY PROTEINS	21
<i>Tat</i>	21
<i>Rev</i>	22
1.2.4 ACCESSORY PROTEINS.....	23
<i>Nef</i>	23
<i>Vif</i>	24
<i>Vpr</i>	24
<i>Vpu</i>	25
1.3 THE HIV-1 LIFE CYCLE.....	26
1.3.2 REVERSE TRANSCRIPTION.....	32
1.3.3 INTEGRATION.....	33
1.3.4 TRANSCRIPTION OF VIRAL DNA	35
1.3.5 ASSEMBLY AND BUDDING	36
1.4 PATHOGENESIS OF HIV-1.....	37
1.4.1 TRANSMISSION OF HIV-1	37
1.4.2 DISEASE PROGRESSION.....	38
1.4.3 VIRAL DYNAMICS AND IMMUNE RESPONSES	39
1.4.4 ANTIVIRAL THERAPY	42
1.5 GENETIC DIVERSITY OF HIV-1	43
1.5.1 SOURCES OF MUTATION DURING HIV REPLICATION.....	44
1.5.2 VIRAL TURNOVER IN HIV INFECTED INDIVIDUALS.....	47
1.5.3 FIXATION RATE OF MUTATIONS IN HIV	49

1.6 RECOMBINATION IN HIV-1	54
1.6.1 MOLECULAR MECHANISMS OF RETROVIRAL RECOMBINATION	55
1.6.2 RECOMBINATION BETWEEN GLOBAL HIV-1 ISOLATES	60
1.6.3 INTRASUBTYPE RECOMBINATION.....	63
1.6.4 RECOMBINATION BETWEEN CLOSELY RELATED HIV-1 STRAINS	64
1.7 MOLECULAR DETERMINANTS OF TROPISM	66
1.7.1 CD4.....	66
1.7.2 CORECEPTORS	68
1.7.3 FURTHER FACTORS INVOLVED IN VIRAL ENTRY.....	74
1.7.4 ENVELOPE DETERMINANTS OF HIV-1 CELLULAR TROPISM AND CORECEPTOR USE	77
1.7.5 CD4 AND CORECEPTOR INDEPENDENT INTERACTIONS	80
1.8 COMPARTMENTALISATION OF HIV-1	83
1.9 HIV-1 AND THE BRAIN	84
1.9.1 NEUROLOGICAL SYMPTOMS AND PATHOLOGY ASSOCIATED WITH HIV-1 INFECTION	85
1.9.2 TIMING OF HIV-1 INFECTION OF THE BRAIN.....	87
1.9.3 INFECTED CELL TYPES IN THE BRAIN	88
1.9.4 HIV-1 INDUCED DAMAGE IN THE CNS	89
1.9.5 HIV-1 NUCLEOTIDE SEQUENCES IN THE BRAIN.....	90
1.9.6 PHENOTYPE OF HIV-1 IN THE BRAIN	92
1.10 MOLECULAR EVOLUTION	97
1.10.1 DISTANCE MEASURES FOR NUCLEOTIDE SEQUENCES	97
1.10.2 MOLECULAR PHYLOGENETICS.....	98
1.10.3 STATISTICAL SIGNIFICANCE OF TREE TOPOLOGIES	100
1.11 AIMS	100
2: MATERIALS AND METHODS	103
2.1 AUTOPSY SAMPLES	104
2.2 BIOLOGICAL SAFETY WHEN WORKING WITH HIV	104
2.3 DNA EXTRACTION	105
2.4 QUANTITATION OF EXTRACTED DNA	106
2.5 POLYMERASE CHAIN REACTION	107
2.5.1 GENERAL PROCEDURES	107
2.5.2 PRIMER SEQUENCES	107
2.5.3 PCR OF p17 ^{GAG} , V1/V2, V3, NEF AND EGFP	109
<i>Reagents</i>	109
<i>Protocol</i>	110
<i>Thermal cycling conditions</i>	110
2.5.4 PCR OF FULL-LENGTH <i>ENV</i> AND LUCIFERASE GENES.....	111
<i>Reagents</i>	112
<i>Protocol</i>	112
<i>Thermal cycling conditions</i>	112
2.6 AGAROSE GEL ELECTROPHORESIS	112
<i>Reagents</i>	112
<i>Protocol</i>	113
2.7 QUANTITATION OF HIV-1 PROVIRAL DNA	113

2.8 CLONING OF PCR PRODUCTS.....	113
2.8.1 PHOSPHORYLATION OF PCR PRODUCTS	114
2.8.2 LIGATION REACTION.....	114
2.8.3 TRANSFORMATION.....	115
<i>Reagents</i>	115
<i>Protocol</i>	116
2.8.4 COLONY SCREENING	116
2.8.5 PLASMID DNA MINI-PREPS.....	117
<i>Reagents</i>	117
<i>Protocol</i>	117
2.8.6 PLASMID DNA MIDI-PREPS.....	118
<i>Protocol</i>	118
2.9 RESTRICTION DIGESTS	119
2.9.1 P _{ME} I DIGEST	119
<i>Reagents</i>	119
<i>Protocol</i>	120
2.9.2 <i>HIND</i> III AND <i>XMN</i> I SINGLE AND DOUBLE DIGESTS:	120
<i>Reagents</i>	120
<i>Protocol</i>	120
2.10 OVERLAP EXTENSION PCR REACTIONS	121
2.11 PCR USING LIGATION REACTIONS AS TEMPLATE.....	121
2.12 EXTENSION REACTIONS	122
2.13 FINAL AMPLIFICATION.....	122
2.14 PCR PRODUCT PURIFICATION.....	123
2.15 DNA SEQUENCING.....	124
2.15.1 MANUAL SEQUENCING	124
<i>Reagents</i>	124
<i>Protocol</i>	125
2.15.2 ACRYLAMIDE GEL ELECTROPHORESIS	126
<i>Reagents</i>	126
<i>Protocol</i>	126
2.15.3 AUTOMATED SEQUENCING	127
<i>Reagents</i>	127
<i>Protocol</i>	127
2.16 SEQUENCE ANALYSIS	128
2.17 EUKARYOTIC CELL CULTURE.....	132
2.17.1 GROWTH MEDIUM AND CELL LINES	132
2.17.2 THAWING CELLS.....	132
2.17.3 PASSAGING CELLS.....	133
2.17.4 FREEZING CELLS	133
2.18 ELECTROPORATION OF 293 T CELLS	133
2.19 FLUORESCENT MICROSCOPY.....	134
2.20 LUCIFERASE ASSAY.....	134
2.21 NEF IMMUNOSTAINING.....	135

2.22 PSEUDOTYPE VIRUS PRODUCTION AND ANALYSIS	136
2.23 CALCIUM PHOSPHATE MEDIATED TRANSFECTION:	138
2.24 VIRAL INFECTIVITY ASSAYS	139
2.25 P24 IMMUNOCYTOCHEMISTRY	140
CHAPTER 3: RESULTS (I)	142
3.1 INTRODUCTION	143
3.2 AIMS	148
3.3 CLINICAL DETAILS OF STUDY SUBJECTS	149
3.4 METHODS	151
3.5 RESULTS	152
3.5.1 V3 AND P17 ^{GAG} TREE TOPOLOGY	152
NA021	152
NA128	155
NA173	155
NA234	160
3.5.2 SEQUENCE DIVERGENCE AND PATHOLOGY APPEARANCE	161
3.6 DISCUSSION	161
3.6.1 TIMING OF VIRAL ENTRY INTO THE CNS.	161
3.6.2 IDENTIFICATION OF IN VIVO RECOMBINATION EVENTS IN THE EVOLUTION OF REGIONAL VIRAL POPULATIONS.....	164
CHAPTER 4: RESULTS (II)	169
4.1 INTRODUCTION	170
4.2 AIMS	175
4.3 MATERIALS AND METHODS	175
4.4 RESULTS	176
4.4.1 NUCLEOTIDE MISINCORPORATION DURING PCR	176
4.4.2 <i>IN VITRO</i> RECOMBINATION DURING AMPLIFICATION BY PCR	177
4.4.3 LIMITING DILUTION <i>ENV</i> PCR	184
4.4.4 GENETIC DIVERSITY COMPARISON BETWEEN LIMITING DILUTION AND CLONING	186
4.4.5 SEQUENCE ANALYSIS OF LIMITING DILUTION <i>ENV</i> SEQUENCES	192
4.4.6 CONSENSUS PSEUDOTYPE VIRUS PRODUCTION	217
4.4.7 PLASMID REARRANGEMENTS DURING EXPANSION IN <i>E. COLI</i>	221
4.5 DISCUSSION	223
CHAPTER 5: RESULTS (III)	244
5.1 INTRODUCTION	245
5.2 AIMS	249
5.3 MATERIALS AND METHODS	249

5.4 RESULTS	249
5.4.1 ASSESSMENT OF LINEAR PCR PRODUCT DNA FOR TRANSIENT GENE EXPRESSION	249
5.4.2 DEVELOPMENT OF PCR EXPRESSION CONSTRUCTS	250
<i>EGFP</i>	254
<i>Luciferase</i>	256
<i>HIV-1 nef</i>	256
<i>HIV-1 env</i>	259
5.5 DISCUSSION	265
CHAPTER 6: FINAL DISCUSSION.....	269
FUTURE WORK	279
BIBLIOGRAPHY	280
APPENDIX.....	330

Declaration

The results presented in this thesis and its composition were undertaken solely by the author unless otherwise stated.

Matthew David Marsden

Edinburgh, 2003

Acknowledgements

I would like to thank the following people:

My principal supervisor Professor Peter Simmonds for providing me with the opportunity to undertake a PhD.

My second supervisor Professor Jeanne Bell for her support and encouragement.

All members of the Viral evolution and HIV groups past and present, in particular Dr Lisa Jarvis, Dr Donald Smith, Dr Ting-Huei Wang, and Dr Sarah McBreen, with a special thanks to Dr Stuart Imlach and Miss Kirsty Newman, for their friendship, support, and endless encouragement.

Dr Paul Clapham, and all the staff and students at the Wohl Virion Centre for providing advice and reagents for the generation of Env-pseudotype virus, and allowing me to perform experiments within their laboratory.

Finally, thanks to all my friends and family.

Abbreviation list

ADC	AIDS dementia complex
AGM	African green monkey
AIDS	Acquired immunodeficiency syndrome
APS	Ammonium persulphate
ARD	AIDS related dementia
ARV	AIDS-associated retrovirus
BBB	Blood-brain-barrier
BGH	Bovine growth hormone
BSA	Bovine serum albumin
C1-C5	Conserved regions of gp120
CA	Capsid protein (p24)
CFTR	Cystic fibrosis transmembrane conductance regulator
CMV IE	Cytomegalovirus immediate early promotor/enhancer
CNS	Central nervous system
cPPT	central PPT
CRF	Circulating recombinant forms
CTLs	Cytotoxic T cells
DAB	Diaminobenzidine
DC-SIGN	Dendritic cell-specific inter cellular adhesion molecule-grabbing nonintegrin
DMEM	Dulbecco's Modified Eagle's Medium
dNTP	deoxynucleoside triphosphates (dATP, dCTP, dITP, dUTP)
DTT	Dithiothreitol
ECL	Extracellular loop
EGFP	Enhanced green fluorescent protein
EIAV	Equine infectious anaemia virus
Env	HIV envelope
FCS	Foetal calf serum

Gag	Group specific antigens
Gal-C	Galactocerebroside
gp120	Surface glycoprotein
gp41	Transmembrane glycoprotein
HAART	Highly active antiretroviral therapy
HAD	HIV associated dementia
HBS	HEPES-Buffered saline
HCV	Hepatitis C virus
HGV/GBV-C	Hepatitis G virus
HIV	Human immunodeficiency virus
HIVE	HIV encephalitis
HTLV	Human T- lymphotropic viruses
ICAM	Intercellular adhesion molecule
IN	Integrase
J-C distances	Jukes and Cantor distances
LAV	Lymphadenopathy-associated virus
LECs	Linear expression constructs
LFA-1	Leukocyte function-associated antigen 1
LN	lymph node
LT	Left temporal
LTR	Long terminal repeats
MA	Matrix protein
mAbs	monoclonal antibodies
MDM	Monocyte derived macrophages
MHC	Major histocompatibility complex
MIP	Macrophage inflammatory protein
M-tropic	Macrophage tropic
MVECs	Microvascular endothelial cells
N	Total population size
NC	p7 nucleocapsid protein

N_e	Effective population size
Nef	Negative factor
N-J	Neighbour-Joining
NNRTIs	Non-nucleoside reverse transcriptase inhibitors
NRTIs	Nucleoside reverse transcriptase inhibitors
NSI	Nonsyncytium-inducing
ORF	Open reading frame
PBMC	Peripheral blood mononuclear cell
PBS	Phosphate Buffered Saline
pbs	primer binding site
PCR	Polymerase chain reaction
PI	Protease inhibitors
PIC	Preintegration complex
Pol	Polymerase
PPT	Polypurine tract
PR	Protease
Pr160 ^{Gag-Pol}	Gag-Pol fusion protein
RANTES	Regulated upon activation normal T cell expressed and secreted
Rev	Regulator of expression of viral proteins
RO	Right occipital lobe
RRE	Rev response element
RT	reverse transcriptase
RT-PCR	Reverse Transcription polymerase chain reaction
SAP	Shrimp alkaline phosphatase
SDF-1	Stromal derived factor-1
SI	Syncytium-inducing
SIV	Simian immunodeficiency virus
TAR	Tat-responsive element
Tat	Transactivator of transcription
TCR	T-cell receptor

<i>Taq</i>	<i>Thermus aquaticus</i>
T-tropic	T-cell tropic
V1-V5	Variable regions of gp120
Vif	Viral infectivity factor
Vpr	Viral protein R
Vpu	Viral protein U

List of Figures

	<i>Page</i>
Figure 1.1	The HIV-1 virion. 9
Figure 1.2	HIV-1 Genome and Encoded Proteins. 11
Figure 1.3	Linear representation of HIV-1 Env glycoprotein. 17
Figure 1.4	Folded representation of HIV-1 gp120. 19
Figure 1.5	The HIV-1 Life Cycle. 28
Figure 1.6	Typical Course of HIV Infection. 40
Figure 2.1.	Schematic representation of primer binding sites on <i>env</i> gene. 127
Figure 2.2.	Vectors used for pseudotype virus production. 137
Figure 3.1	Phylogenetic analysis of p17 ^{gag} and V3 sequences from subject NA021. 153
Figure 3.2	Phylogenetic analysis of p17 ^{gag} and V3 sequences from subject NA128. 156
Figure 3.3	Phylogenetic analysis of p17 ^{gag} and V3 sequences from subject NA173. 157
Figure 3.4	Phylogenetic analysis of p17 ^{gag} and V3 sequences from subject NA234. 158
Figure 3.5	Comparison of inferred V3 region amino acid sequences of variants from study subjects. 159
Figure 4.1	Schematic representation of mutations that occurred during <i>env</i> PCR using cloned viral templates. 178
Figure 4.2	System for detecting recombination during <i>env</i> PCR amplification. 179
Figure 4.3	Detection of recombinant <i>env</i> PCR products. 181
Figure 4.4	Phylogenetic trees constructed with V1/V2 sequences obtained by cloning or limiting dilution PCR. 188
Figure 4.5	V1/V2 nucleotide and predicted amino acid sequences obtained either by cloning or by limiting dilution PCR 190

Figure 4.6	Example of <i>env</i> PCR products amplified from single proviral templates	194
Figure 4.7	Predicted Env (gp160) amino acid sequence of PCR products amplified at limiting dilution from different tissue samples obtained at autopsy.	195
Figure 4.8	Predicted amino acid sequences of the second exons of Tat and Rev, encoded within the gp160 open reading frame.	203
Figure 4.9	Phylogenetic tree constructed with limiting dilution <i>env</i> (gp160) nucleotide sequences.	205
Figure 4.10	Group-scan plots for right occipital limiting dilution <i>env</i> sequences.	208
Figure 4.11	Group-scan plots for left temporal limiting dilution <i>env</i> sequences	210
Figure 4.12	Group-scan plots for lymph node limiting dilution <i>env</i> sequences	211
Figure 4.13	Comparison of group-scan with bootscan	215
Figure 4.14	Schematic representation of consensus clone and pseudotype virus production.	216
Figure 4.15	Example of <i>env</i> clone instability in <i>E. coli</i>	222
Figure 5.1	Expression of PCR product in 293T cells.	251
Figure 5.2	Schematic representation of amplification steps used in production of PCR expression constructs	253
Figure 5.3	EGFP PCR expression construct	255
Figure 5.4	Luciferase PCR expression construct.	257
Figure 5.5	Nef expression	260
Figure 5.6	Env expression	261

List of Tables

	<i>Page</i>
Table 1.1	71
Coreceptors that support primate lentivirus infection of CD4+ cell lines <i>in vitro</i> .	
Table 2.1.	108
Oligonucleotide primer sequences and 5' nucleotide binding sites on template DNA	
Table 2.2.	111
Primers pairs used in PCR and sequencing reactions.	
Table 2.3.	115
Composition of ligation reaction.	
Table 2.4.	122
Primer pairs used in construction of linear expression constructs.	
Table 2.5.	132
Mammalian cell lines used for transfection experiments, HIV-1 receptors expressed and selective media required for maintenance of expression	
Table 3.1	150
Clinical background and pathology appearance of study subjects.	
Table 3.2	150
p17 ^{gag} and V3 sequence divergence.	
Table 4.1	185
Sensitivity comparison of env PCR with V3 PCR using cloned HIV proviral DNA as template.	
Table 4.2	185
Sensitivity comparison of env PCR with V3 PCR using HIV proviral DNA extracted from autopsy tissue as template.	
Table 4.3	220
Result of pseudotype single round infection assay using consensus <i>env</i> clones.	
Table 5.1	264
Infectious titre of pseudotype virus on NP2 cells expressing CD4 and CCR5.	

Abstract

Infection with human immunodeficiency virus (HIV) is associated with a slow deterioration of the immune system culminating in acquired immunodeficiency syndrome (AIDS). Approximately 42 million people worldwide are currently living with HIV/AIDS and new infections are occurring at a rate of around 5 million per year. A common consequence of HIV type 1 (HIV-1) infection is the development of neurological symptoms, which often manifest as a syndrome of cognitive and motor dysfunction termed HIV-associated dementia (HAD). HIV-1 is frequently detected within the brain of infected individuals, where the major reservoir of replicating virus is monocyte-derived macrophages and microglial cells. The resultant HIV encephalitis (HIVE) is often characterised by the presence of syncytia, which are a product of HIV-1 envelope (Env) protein-directed fusion of infected cells. Further HIV-induced damage to neuronal cell types can occur by a number of different mechanisms, and may directly contribute to the neurological symptoms experienced by infected individuals. Nucleotide sequence analysis of HIV-1 variants in the brain of individuals with HIVE, usually based upon partial *env* gene sequences, often demonstrates the presence of a viral subpopulation that is genetically distinct from viral variants present in the lymphoid system. Several different factors could lead to such sequence segregations, including viral adaptation, physical isolation of the different subpopulations, and variation in virus turnover rates in different anatomical compartments. Determining what underlies these frequently observed genetic divisions is important for a more complete understanding of HIV-1 neuropathogenesis. Thus the aim of this project was to further characterise the

evolutionary relationships between HIV-1 variants in the brain and lymphoid system, and in particular to search for evidence of recombination between variants in the two compartments by analysing sequence segments both within and outside the major genetic determinants of HIV-1 cellular tropism.

The evolutionary relationships manifest in the p17^{gag} and V3 genomic regions were assessed between HIV-1 variants derived from the brain and lymphoid system of four infected individuals with evidence of HIV. p17^{gag} genetic distances between brain and lymph node variants were calculated, allowing approximate divergence times to be inferred. In two cases this indicated that p17^{gag} sequences present in the brain and lymphoid compartments had diverged prior to the individual progressed to AIDS, possibly indicating an autonomous pre-AIDS infection of the brain in these two study subjects. A correlation was also noted between the severity of HIV and both the V3 sequence divergence within brain, and the extent to which variants in brain differed from those in the lymphoid system. In addition, p17^{gag} and V3 phylogenetic tree topologies generated with sequences obtained from the same tissues exhibited discordant phylogenetic relationships in each study subject. The most obvious explanation for these complex evolutionary relationships is the occurrence of recombination among HIV-1 variants present within the brain, and also between brain and lymphoid derived HIV-1 variants.

In order to characterise this *in vivo* recombination more comprehensively it was necessary to obtain longer sequences that may contain identifiable recombination breakpoints between viral variants from different tissues. The entire HIV-1 *env* gene was

chosen for this application. To ensure rigorous sequence analysis it was first essential to quantify and eliminate a range of potentially confounding *in vitro* artefacts associated with the polymerase chain reaction (PCR) amplification and subsequent molecular cloning of HIV-1 *env* sequences. Artefacts included in this assessment were the incorporation of single nucleotide errors during PCR, recombination during PCR and resampling of clones. The results demonstrated that each of these artefacts could significantly affect the final sequences obtained. In particular, recombination during PCR was found to occur at a very high level even when small numbers of template molecules were present at the start of the reaction. Resampling of clones was also evident despite addition of a large number of amplifiable template molecules to the PCR. Limiting dilution PCR was therefore used to obtain entire HIV-1 *env* sequences for genetic analysis. Inter-tissue recombination was apparent in several *env* sequences, providing further evidence for the occurrence of this mutational mechanism in the context of a single HIV-1 infection. Close relationships between particular brain and lymphoid tissue derived HIV-1 variants were also noted in certain genomic regions, despite apparent tissue specificity in others.

If artefactually altered sequences are subsequently used in phenotypic assays, this may lead to alterations in the observed functional properties of encoded proteins. These considerations are especially important in the case of HIV-1 Env, where receptor interactions can be both conformationally complex and strain dependent, and the potential determinants of properties such as neurotropism and neurovirulence are poorly defined. Therefore two approaches were designed with the aim of maintaining the exact

characteristics of the *env* sequences present *in vivo*. The first involved consensus cloning, whereby *env* sequences amplified from single proviral templates were cloned and ten clones mixed before use in an Env-pseudotype virus assay. This approach resulted in the production of infectious virus, which in principal would be expected to contain no errors due to *in vitro* processes. However, issues with clone stability during plasmid expansion in *E. coli* restricted its general utility and also raised the possibility that defective envelopes may be preferentially isolated during the cloning process. Therefore in the second approach a novel technique was developed for the addition of promoter and polyadenylation sequences directly to PCR products, allowing expression of open reading frames in mammalian cells without the need for prior molecular cloning. The efficiency of this technique was verified by using two reporter genes and HIV-1 *env* and *nef* sequences derived from autopsy tissue. As envelope genes were amplified at limiting dilution, use of this method ensured that sequence characteristics and genetic linkages present *in vivo* were maintained, and not affected by the various *in vitro* artefacts that may significantly alter retrieved sequences.

Together these results indicate that recombination between HIV-1 variants that have diverged during a single infection is an extremely common event. The identification of recombination between brain and lymphoid derived virus also suggests that physical separation or differences in viral turnover rates are not the source of the commonly observed genetic differences between viral variants within these two compartments. Therefore it may be speculated that those regions of the viral genome that do show tissue specificity are the result of an adaptive process.

Chapter 1: General Introduction

1.1 Historical perspective

The acquired immunodeficiency syndrome (AIDS) first came to public notice in 1981, when five unusual cases of young, previously healthy homosexual men with *Pneumocystis carinii* pneumonia were registered in the USA (CDC, 1981). It soon became evident that members of other groups including intravenous drug users, hemophiliacs, blood transfusion recipients, and recent Haitian immigrants to the USA were also being affected by what appeared to be a new immunodeficiency disease. This unusual syndrome was characterised by the presence of opportunistic infections (*Pneumocystis carinii* pneumonia, *Toxoplasma gondii* encephalitis, cytomegalovirus-associated retinitis, and cryptococcal meningitis), generalised lymphadenopathy, and a range of unusual cancers (non-Hodgkin's lymphoma and Kaposi's sarcoma). Significantly, the sexual partners and children of members of the various risk groups often presented with similar symptoms. Together these epidemiological observations suggested that the disease was caused by a novel pathogen, which was being transmitted within contaminated blood or via sexual intercourse with an affected individual (Marmor et al., 1982; Jaffe et al., 1983; Cowan et al., 1984).

In 1983 Barré-Sinoussi and colleagues provided the first evidence that a retrovirus was the etiological agent of AIDS. These researchers recovered a virus containing reverse transcriptase (RT) from the lymph node of an individual with lymphadenopathy, and termed it lymphadenopathy-associated virus (LAV) (Barre Sinoussi et al., 1983). Subsequently, a group led by Gallo also isolated retroviruses from AIDS patients. Believing them to be closely related to the human T-

lymphotropic viruses (HTLV), and to distinguish them from the noncytopathic HTLV-I, and HTLV-II, these researchers named the new agent HTLV-III (Gallo et al., 1984). At the same time Levy and colleagues isolated a similar retrovirus from both AIDS patients and healthy individuals from the various risk groups, and named the virus AIDS-associated retrovirus (ARV) (Levy et al., 1984). It became apparent that LAV, HTLV-III, and ARV shared many common properties and in 1986 the International Committee for the Taxonomy of Viruses ultimately named the causative agent of AIDS the human immunodeficiency virus (HIV), and later HIV type 1 (HIV-1) (Coffin et al., 1986). In 1986 a related but genetically quite distinct and less pathogenic retrovirus was isolated from individuals in western Africa (Barin et al., 1985; Clavel et al., 1986a; Clavel et al., 1986b). This virus was termed HIV type 2 (HIV-2).

The World Health Organisation estimated that at the end of 2002 approximately 42 million people worldwide were living with HIV/AIDS. Around five million new HIV infections and 3.1 million AIDS deaths occurred in that year alone. The best current projections suggest that unless a drastically expanded prevention effort is mounted, then a further 45 million people will become infected between the years 2002 and 2010. HIV prevalence is currently highest in Sub-Saharan Africa, with an estimated 29.4 million people infected. Rapidly expanding epidemics are also evident in areas of Asia and the Pacific. Approximately 980 thousand HIV infected individuals are currently living in North America and the corresponding figure for Western Europe is estimated at 570 thousand.

1.1.1 Classification and morphology of retroviruses

HIV is classified as a member of the *Lentivirus* genus in the family *Retroviridae*. Two features of the retroviral life cycle (when together) distinguish retroviruses from members of all other virus families: Reverse transcription of viral genomic RNA into DNA, and integration of the viral DNA genome into the host cell chromosome. Based upon features such as genome structure and phylogenetic relationships, retroviruses have been classified into 7 different genera. These are the alpharetroviruses (e.g. Rous sarcoma virus), betaretroviruses (e.g. mouse mammary tumor virus), gammaretroviruses (e.g. murine leukemia virus), deltaretroviruses (e.g. human T-lymphotropic virus 1), epsilonretroviruses (e.g. walleye dermal sarcoma virus), lentiviruses (e.g. human immunodeficiency virus type 1), and spumaviruses (e.g. human foamy virus). Alpha- beta- and gamma-retroviruses are considered “simple” and encode only group specific antigens (Gag), protease (PR), polymerase (Pol), and envelope (Env) gene products. Members of the remaining retroviral genera are considered “complex”, and also encode a number of regulatory proteins with diverse functions (Goff, 2001).

The name *Lentivirus* is derived from *lentus*, Latin for slow. Lentiviral infections are characterised by persistent viral replication, long and variable incubation periods, neurological manifestations, and the destruction of immunologic or hematologic cells (Desrosiers and Letvin, 1987). Common features of lentiviruses include the presence of regulatory genes not found in other retroviruses, similar virion morphology, extensive genetic and antigenic diversity, and a tropism for macrophages. The *Lentivirus* genus may be split into 6 groups by virtue of primary host species and other features: primates (e.g. HIV, simian immunodeficiency virus), horses (e.g.

equine infectious anaemia virus; EIAV), cats (e.g. feline immunodeficiency virus), sheep (e.g. Visna virus), goats (e.g. caprine arthritis-encephalitis virus), and cattle (e.g. bovine immunodeficiency virus) (Coffin, 1996). The first account of what later transpired to be a lentiviral disease was equine infectious anemia in 1904 (Vallee and Carre, 1904). Subsequently the description of Visna (a neurological disease in sheep caused by a lentivirus), gave rise to the concept of slow disease onset that is characteristic of many lentiviral infections (Sigurdsson, 1954).

In 1971 the human foamy virus was described, representing the first reported retroviral infection in humans (Achong et al., 1971). More recently, interest in retroviruses has greatly expanded. This primarily reflects their role as human and animal pathogens, but other retroviral properties have proved valuable in diverse areas of biological research. For example retroviruses have been of great use in gene therapy applications, as evolutionary markers, in the identification of proto-oncogenes, and as a source of enzymatic tools for molecular biology.

Mature retrovirus virions are slightly pleomorphic spheres of approximately 100 nm diameter. The virions are enveloped in a lipid bilayer containing envelope glycoprotein spikes. Within the envelope is a roughly spherical shell consisting of matrix proteins. Retroviral cores may be spherical, conical, or rod shaped, and are largely composed of capsid proteins. Two copies of highly condensed genomic RNA are associated with nucleocapsid protein and carried within the core. Viral enzymes and other components required for replication, which can include accessory gene products, are also present in the virion. The genome is a dimer of linear, positive

sense, single stranded RNA of between approximately 7 Kb and 12.3 Kb in length (Vogt, 1997).

1.1.2 Origins and Global Diversity of HIV

HIV-1 and HIV-2 each originate from the zoonotic transmission of simian immunodeficiency virus (SIV) into humans from different primate sources.

HIV-1 is closely related to SIV_{CPZ}, a virus endemic to *Pan troglodytes troglodytes* (chimpanzees) in Central Africa (Gao et al., 1999). Phylogenetic analysis of diverse HIV-1 strains has revealed the presence of three clearly distinct clades of viruses, termed groups M (main), N (new, or non-M, non-O), and O (outlier). These lineages are interspersed with SIV_{CPZ} lineages, and therefore must have arisen from separate cross-species transmission events (Hahn et al., 2000; Gao et al., 1999; Corbet et al., 2000). Group O strains are generally localised in West Central Africa (Deleys et al., 1990; Peeters et al., 1997; Zekeng et al., 1994), and group N strains are only represented by a few isolates from Cameroon (Simon et al., 1998). HIV-1 group M strains have spread throughout the world and are responsible for the great majority of global HIV infections. Group M viruses fall into phylogenetically distinct clades, and there are currently 9 recognised group M subtypes designated A, B, C, D, F, G, H, J, and K (Robertson et al., 2000). In Env proteins approximately 25% to 35% amino acid sequence differences are exhibited between subtypes, and up to 20% differences within subtypes (Sharp et al., 1999). Group M subtypes vary in prevalence in different geographical areas. Representatives of all subtypes are present in Africa, and subtype C has the greatest overall worldwide prevalence. Subtype B strains are responsible for the majority of infections in North America and Western Europe

(Thomson et al., 2002). A large number of inter-subtype recombinants have also arisen, and in some geographical areas are responsible for the majority of HIV infections. For example an A/G recombinant is very common in parts of West and Central Africa (McCutchan et al., 1999), and an A/E recombinant has a high prevalence in many parts of Asia (Gao et al., 1996b).

HIV-2 was originally identified in West Africa (Barin et al, 1985; Clavel et al, 1986a; Clavel et al, 1986b) and is closely related to SIV_{smm} in *Cercocebus atys* (sooty mangabeys). Therefore this primate species is believed to be the source of HIV-2 in humans (Gao et al., 1992; Hirsch et al., 1989). HIV-2 is less pathogenic than HIV-1, and infection with HIV-2 is associated with an increased time between initial infection and the diagnosis of AIDS (Marlink et al., 1994). HIV-2 is also less readily transmitted than HIV-1 (Kanki et al., 1994), perhaps explaining why it has not spread significantly from West Africa where HIV-2 prevalence rates are stable or even decreasing (Schim van der Loeff MF and Aaby, 1999).

With regard to the timing of HIV entry into human populations, the earliest direct evidence for HIV infection in humans comes from the retrospective analysis of an African blood sample stored from 1959 (Zhu et al., 1998). The recovered sequence was phylogenetically positioned near the ancestral nodes of HIV-1 group M subtypes B and D, and close to the centre of the group M radiation. This indicated that HIV-1 group M originated from a single cross-species transmission event, which occurred not long before 1959 (Zhu et al, 1998). More recently this approximation has been refined by the use of sequence data from a large number of viral isolates to estimate the coalescence time of group M (Korber et al., 2000). The results of this analysis

suggest that the cross-species transmission event occurred between the years 1915 and 1941, most likely around 1931 (Korber et al, 2000). Currently, the most accepted theory for the method of HIV introduction into humans is through acquisition of SIV_{CPZ} by contact with contaminated chimpanzee meat or blood during the acquirement, processing, or consumption of bushmeat.

1.2 The HIV-1 Virion

Mature HIV-1 virions are approximately 100 nm in diameter, with a cone shaped core composed of p24 capsid (CA) protein (Fig. 1.1). Within this core are two copies of the positive sense RNA genome, closely associated with the p7 nucleocapsid protein (NC). Also within the core are the reverse transcriptase (RT), protease (PR), integrase (IN), Vpr, and p6 proteins. Small amounts of the Nef and Vif viral proteins are incorporated into virions, but this incorporation may be non-specific and its significance is unclear. Host-cell derived cyclophilin A is incorporated through a specific association with p24. The virion is enveloped by a lipid bilayer derived from the host cell membrane, and contains host cell cytoplasmic membrane-associated proteins in addition to the virally encoded gp41 transmembrane (TM) and gp120 surface (SU) envelope glycoproteins. gp41 and gp120 are non-covalently associated as heterodimers, and arranged as trimers on the surface of the virion. The p17 matrix protein (MA) is closely associated with the inner side of the viral membrane (Freed and Martin, 2001).

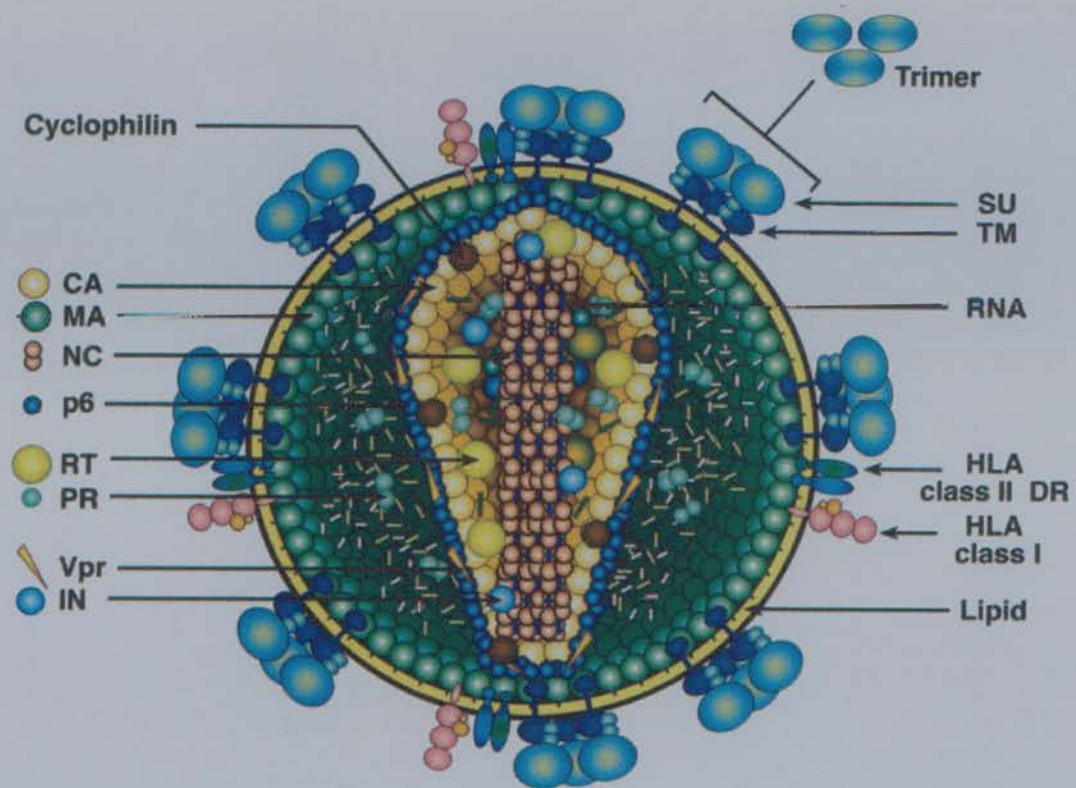


Figure 1.1. The HIV-1 virion. Schematic representation of a cross section through the mature HIV-1 virion. The host cell-derived lipid bilayer contains trimeric Env glycoproteins composed of SU (gp120) and TM (gp41) heterodimers. Host cell membrane proteins such as HLA molecules are also incorporated during virion budding. The MA protein is closely associated with the inner surface of the lipid membrane. CA forms the viral core, and specifically incorporates the host protein cyclophilin into the virion. Two copies of single stranded genomic RNA coated in NC protein are located within the core. The three viral enzymes PR, RT, and IN are also present within the virion, as are the Vpr and p6 viral proteins. Figure copied from (Vogt, 1997).

1.2.1 Genomic organisation of HIV-1

HIV-1 is a complex retrovirus and therefore includes a number of accessory genes in addition to the *gag*, *pol*, and *env* open reading frames (ORFs) that are common to all retroviruses. The 9.2 kilobase (Kb) HIV-1 RNA genome contains nine ORFs and is flanked by regulatory sequences that form components of the proviral long terminal repeats (LTRs). HIV-1 genomic RNA also contains a number of *cis*-acting elements, which are necessary for efficient and coordinated viral replication. A diagram of the HIV-1 genome and encoded proteins is shown (Fig. 1.2). The *gag*, *pol*, and *env* genes encode structural proteins and are translated as polyprotein precursors (Pr55^{Gag}, Pr160^{Gag-Pol}, and gp160 respectively), which are subsequently cleaved into their component proteins by viral or cellular proteases. *gag* encodes the matrix (MA, p17), capsid (CA, p24), nucleocapsid (NC, p7), and p6 proteins. *pol* encodes the three viral enzymes, protease (PR, p10), the heterodimeric reverse transcriptase (RT), and integrase (IN, p32). *env* encodes for surface (SU, gp120) and transmembrane (TM, gp41) envelope glycoproteins. The remaining six ORFs encode the regulatory and accessory proteins Vif, Vpr, Vpu, Tat, Rev, and Nef (Freed and Martin, 2001).

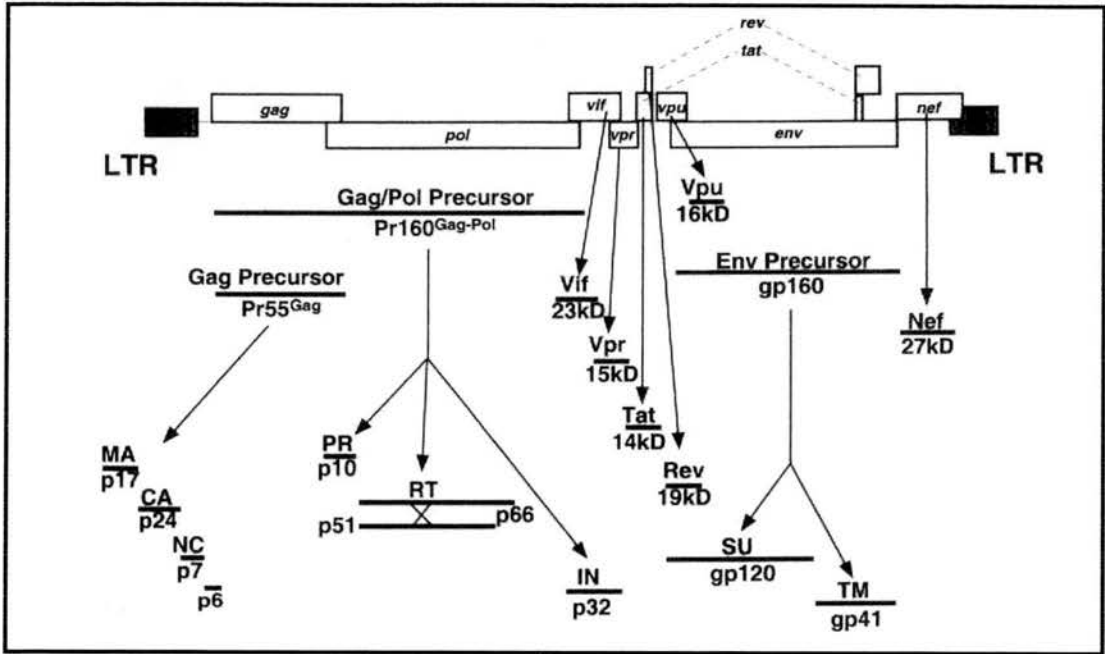


Figure 1.2. HIV-1 Genome and Encoded Proteins. A genome map showing the location and reading frames of the nine HIV-1 genes. Sizes of the primary translation products (in some cases polyproteins), and the processed mature viral proteins are indicated. Figure copied from (Goff, 2001).

1.2.2 Structural proteins

Gag

The 55 kilodalton (KDa) Gag precursor protein (Pr55^{Gag}) is translated from an unspliced 9.2 Kb mRNA transcript. After translation Pr55^{Gag} oligomerises and is targeted to the cellular membrane, where it is cleaved by the viral protease concomitantly with budding to release the MA, CA, NC, and p6 proteins (Gottlinger et al., 1989).

The N-terminus of Pr55^{Gag} comprises the 132 amino acid p17 (MA) protein. MA attaches to the inner surface of the viral membrane and forms a matrix between the envelope and capsid. A host cell N-myristyl transferase post-translationally modifies MA by adding a myristic acid moiety to the N-terminal glycine. This myristate anchors the Pr55^{Gag} polyprotein to the cellular membrane during virion assembly (Ono and Freed, 1999). Additional regions of MA are required for targeting of Pr55^{Gag} to the plasma membrane during this process (Ono et al., 2000; Ono et al., 1997). MA also interacts with gp41 and has a role in the incorporation of Env glycoproteins into the nascent virion during budding (Freed and Martin, 1996; Murakami and Freed, 2000). It has been reported that mutations in MA can lead to impairment of viral DNA synthesis upon infection of a new cell, possibly by causing defects in uncoating (Casella et al., 1997; Kiernan et al., 1998). Other reports suggest that the basic domain of MA has a role in translocation of the preintegration complex into the nucleus (Bukrinsky et al., 1993a; Vonschwedler et al., 1994), and that C-terminal phosphorylation of MA may be important for productive infection of terminally differentiated cells (Gallay et al., 1995).

Positioned next to MA in the Pr55^{Gag} polyprotein is p24 (CA), which forms the viral core. CA is composed of two domains, an N-terminal core domain and a C-terminal dimerisation domain. The core domain is important for correct viral core condensation/maturation, and contains a cyclophilin A-binding loop (Gamble et al., 1996). Incorporation of cyclophilin A into virions enhances infectivity, perhaps by aiding correct CA refolding upon maturation or facilitating the uncoating process in the new host cell, although its specific role in these processes role is unclear (Gamble et al., 1997; Grattinger et al., 1999; Wiegers et al., 1999). The C-terminal dimerisation domain is involved in virion assembly, CA dimerisation, and Gag multimerisation (Gamble et al, 1997). Within the dimerisation domain is a 20 amino acid region known as the major homology region (MHR), which is highly conserved between retroviral genera. The MHR has been implicated in the incorporation of Pr160^{Gag-Pol} into virions and is required for particle formation (Srinivasakumar et al., 1995).

The p7 (NC) protein is associated with genomic RNA within the virion core. NC has two zinc finger motifs, which are essential for its function (Mizuno et al., 1996; Poon et al., 1996). The NC protein binds specifically to an approximately 120 nucleotide sequence of genomic RNA situated at the 5' end of the viral genome around the major splice donor site, known as the packaging signal (ψ) (DeGuzman et al., 1998; Laughrea et al., 1997). Full-length genomic RNA is then delivered by NC to the assembling virion (Clever and Parslow, 1997). NC also performs a variety of other functions in the virus life cycle. These include Gag multimerisation, annealing of tRNA^{Lys3} to the primer binding site, stabilisation of genomic RNA dimers, facilitating initiation and obligatory strand transfers during reverse transcription, and

melting RNA secondary structures (Bennett et al., 1993; Rist and Marino, 2002; Bernacchi et al., 2002; Li et al., 1996; Muriaux et al., 1996).

The 51 amino acid p6 protein is liberated from the C-terminus of Pr55^{Gag}. One function of p6 is to facilitate a step late in virus assembly. More specifically a highly conserved four amino acid motif towards the N-terminus of p6 has been identified as necessary for efficient virion budding (Huang et al., 1995). A further role of p6 is incorporation of Vpr into virions. This function is dependent upon the presence of several conserved residues near the C-terminus, which interact directly with an alpha helical structure at the N-terminus of Vpr (Selig et al., 1999; Kondo and Gottlinger, 1996; Paxton et al., 1993; Bachand et al., 1999).

Pol

The *pol* gene is translated as part of a 160 KDa Gag-Pol fusion protein (Pr160^{Gag-Pol}). This fusion protein is generated by a -1 ribosomal frameshift during translation of *gag*, mediated by a heptanucleotide “slippy sequence” and a short stem loop. The requirement for a frameshift ensures that *pol*-derived proteins are expressed at between 5% and 10% of Gag protein levels. Pr160^{Gag-Pol} is myristylated at the N-terminus and targeted to the plasma membrane. The virally encoded PR cleaves Pr160^{Gag-Pol} at a number of sites, thereby liberating Gag and the three Pol proteins p10 (PR), p66/p51 (RT), and p32 (IN).

During or soon after viral budding, PR first liberates itself from the Gag-Pol precursor protein. A number of sites in both Gag and Gag-Pol are then cleaved by PR, which releases their component proteins and gives rise to the conformational changes associated with virus maturation. PR is an aspartic protease, with two 10

KDa monomers of 99 amino acids forming the functional homodimeric enzyme (Copeland and Oroszlan, 1988; Debouck et al., 1987). The substrate binding site is located at the interface between the two monomers, and each contributes an aspartic acid residue to this active site. Upon substrate binding flexible flaps within the dimer close upon the substrate, creating hydrophobic pockets. The two opposed aspartic acid residues then coordinate a water molecule to catalyse the hydrolysis of a peptide bond in the target protein. Protease activity is also dependent on the presence of the NC and p6 domains, which facilitate dimerisation of the Pr160^{Gag-Pol} precursor protein (Zybarth and Carter, 1995)

The mature RT enzyme is a heterodimeric protein composed of p51 and p66 subunits. p51 is obtained by proteolytic cleavage of the C-terminal 15 KDa RNase H domain from p66, in a reaction performed by the viral PR. Despite being derived from polyproteins with the same amino acid sequence, the p51 and p66 subunits have substantially different folding patterns (Huang et al., 1998). As is the case with many other nucleic acid polymerases, each subunit is composed of four subdomains named fingers, palm, thumb, and connection. An RNase H domain is also present in the p66 subunit. The active site is situated in the palm of p66, and contains three aspartic acid residues that are critical for polymerisation activity. Reverse transcriptase catalyses three distinct enzymatic reactions, RNA-dependent DNA polymerisation, DNA-dependent DNA polymerisation, and degradation of RNA in DNA-RNA heteroduplexes (Jacobomolina et al., 1993; Baltimore, 1970; Temin and Mizutani, 1970). HIV-1 reverse transcription is initiated at the 3' end of a tRNA^{Lys3} primer, which is annealed to the primer binding site (pbs) situated upstream of the *gag* initiation codon. During polymerisation the finger domains close upon the template

and dNTP. This facilitates a nucleophilic attack by the 3'-OH of the primer upon the incoming dNTP. The finger domains may then adopt a more open conformation, allowing release of the pyrophosphate and introduction of the next dNTP into the active site (Huang et al, 1998).

The IN enzyme is a 32 KDa protein whose role is to integrate the double stranded DNA viral genome into the host cell chromosome. IN functions as a multimer, and has 3'-end processing and strand transfer activities (Sayasith et al., 2000). Each monomer is composed of three domains, an N-terminal domain, a core domain, and a C-terminal domain. The N-terminal domain has a zinc finger motif and contributes to multimerisation (Zheng et al., 1996). The core domain plays an essential role in catalysis (Bushman et al., 1993). The C-terminal domain is required for both 3' end processing and integration reactions. Regions of the C-terminus may also facilitate the association of IN with the viral genome during reverse transcription, and nuclear import of the pre integration complex (Gallay et al., 1997; Engelman et al., 1994).

Env

HIV Env proteins are responsible for the binding of virions to CD4 and a coreceptor present on target cells, and the subsequent fusion of viral and host cell membranes.

The *env* gene is translated from singly spliced 4.3 Kb Vpu/Env bicistronic mRNA, by ribosomes associated with the rough endoplasmic reticulum. A linear representation of the Env glycoprotein with structural and functional domains labelled is shown (Fig. 1.3).

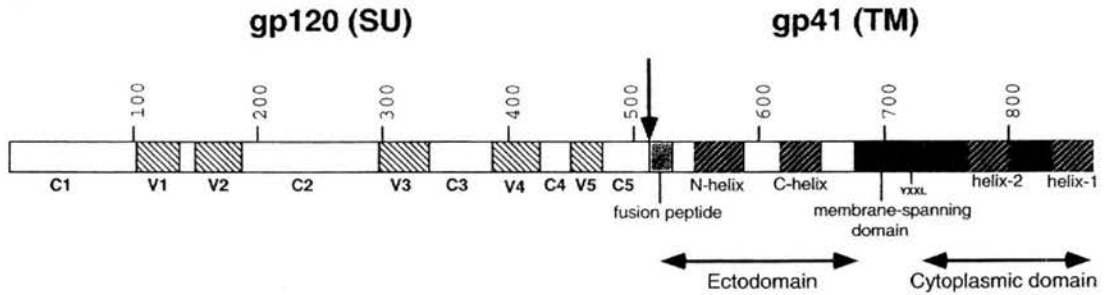


Figure 1.3. Linear representation of HIV-1 Env glycoprotein. The upper arrow indicates the site of gp160 cleavage into gp120 and gp41. In gp120 crossed-hatched areas represent variable domains (V1-V5) and open boxes depict conserved sequences (C1-C5). In the gp41 ectodomain the N-terminal fusion peptide and both N- and C- helices are indicated. The membrane-spanning domain, endocytosis motif (YXXL), and two predicted cytoplasmic helical domains (helix-1 and -2) are also shown. Amino acid numbers are indicated. Figure copied from (Goff, 2001).

Env is synthesised as a 160 KDa polyprotein precursor (gp160) of approximately 845-870 amino acids, which is cotranslationally glycosylated by the addition of asparagine-linked, high mannose sugar chains (Earl et al., 1991; Earl et al., 1990; Chan et al., 1997).

The glycosylation of Env and subsequent post-translational addition of complex sugars is substantial, and it is estimated that half the molecular mass of virion-associated gp120 is composed of oligosaccharide side chains. After translation gp160 oligomerises and is transported to the Golgi apparatus, where it is cleaved into the SU (gp120) and TM (gp41) subunits by a host furin or furin-like protease (Willey et al., 1988; Veronese et al., 1985; Hallenberger et al., 1992). Newly liberated gp41 and gp120 subunits then form a noncovalent association and are transported to the cell surface. This gp41-gp120 interaction is fairly weak, leading to substantial shedding of gp120 from the surface of Env expressing cells. The Env glycoprotein complex is then incorporated into budding virion particles, possibly aided by interactions between the long gp41 cytoplasmic tail and MA (Murakami and Freed, 2000). Complete, virion-associated gp41/gp120 heterodimers are arranged as trimers protruding from the viral envelope (Weiss et al., 1990; Earl et al., 1990).

The surface protein gp120 is composed of five variable regions (V1-V5) interspersed with five more conserved regions (C1-C5) (Starcich et al., 1986). Formation of intramolecular disulphide bonds in gp120 results in the first four variable regions being arranged into looplike structures (Leonard et al., 1990) that are well exposed on the surface of gp120 (Wyatt et al., 1993; Moore et al., 1994). A schematic representation of the gp120 folding pattern is shown (Fig. 1.4).

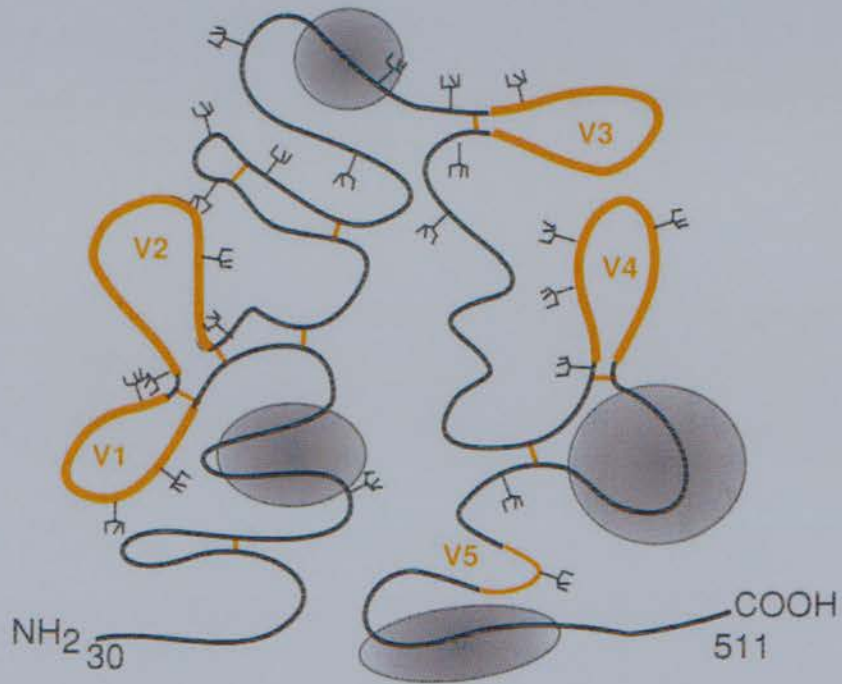


Figure 1.4. Folded representation of HIV-1 gp120. Disulphide bonds and variable regions are shown in colour, N-linked oligosaccharide side chains are shown as branched structures. Grey regions depict domains involved in CD4 binding. Figure copied from (Hunter, 1997).

The conserved regions fold into the gp120 core, which is composed of an inner domain, an outer domain, and a β sheet also referred to as the “bridging sheet” (Wyatt et al., 1998). The inner domain faces the trimer axis and gp41, and the outer domain is exposed on the surface of the trimer. Areas of both domains and the bridging sheet mediate CD4 binding. Following CD4 binding a conformational change occurs in gp120 that forms or exposes the coreceptor binding site (Sattentau et al., 1993; Trkola et al., 1996). Binding of CD4 and the subsequent engagement of a coreceptor molecule are likely to promote other conformational changes in the envelope glycoproteins, which result in activation of gp41 and membrane fusion. The gp120 glycoprotein also contains the principal determinants of HIV-1 cellular tropism. Mutations in gp120 that lead to altered viral phenotypic properties such as second receptor usage and cellular tropism will be discussed further (section 1.7.4).

The transmembrane protein gp41 is responsible for fusion of the viral and host cell membranes. gp41 is similar in many respects to other proteins involved in membrane fusion such as the HA₂ hemagglutinin protein of orthomyxoviruses, the F protein of paramyxoviruses, and the SNAREs and other eukaryotic proteins involved in fusion of intracellular vesicles (Chan et al, 1997; Tan et al., 1997; Dutch et al., 2000; Tamm and Han, 2000; Martin et al., 1999; Jahn and Sudhof, 1999). The N-terminus of gp41, referred to as the fusion peptide, is highly hydrophobic and essential for membrane fusion. Two helical structures are C-terminal to the fusion domain, the N- and C-helices. Prior to membrane fusion these two helices may be kept apart, and held in position by regions of gp120 (Doms and Moore, 2000). In the fusion-activated conformation the two helices of each gp41 molecule in the trimer are arranged in an antiparallel fashion to form six helical bundles (Chan et al, 1997). The

N-helices, each containing a leucine zipper domain, are situated in the centre to form a coiled-coil structure. Conformational changes in gp120 and gp41 that immediately precede membrane fusion have not been fully elucidated. One proposed mechanism is the formation of an extended coiled-coil structure, which facilitates insertion of the fusion peptide into the host cell membrane. Complexing of the N- and C-helices would then bring the two membranes into close proximity. This “spring-loaded” model for membrane fusion is similar to that proposed for influenza virus (Bullough et al., 1994).

1.2.3 Regulatory proteins

Tat (transactivator of transcription)

Tat is a 101 amino acid protein encoded by two exons. The first exon is located just upstream of *env*, and the second overlaps with both the *env* and *rev* ORFs (Arya et al., 1985). A further 72 amino acid “one exon” Tat protein is also produced by translation of only the first exon, and appears to retain all the transactivating properties of the full length protein. When Tat is absent the RNA polymerase II transcriptional complex recruited by promoter and regulatory elements within the LTR is inefficient, and typically results in the transcription of only a few hundred nucleotides (Feinberg et al., 1991). This basal level of transcription is sufficient to allow the expression of small amounts of Tat protein. Tat is a nuclear protein, and its function is to direct the formation of a more processive transcription complex, thereby increasing the levels of viral mRNA by several hundred fold (Alonso and Peterlin, 1999). The activity of Tat is dependent upon binding to the Tat-responsive element (TAR), a 59 nucleotide long stem loop present at the 5' end of all nascent viral RNAs (Feng and Holland, 1988). Tat-defective HIV-1 mutants do not produce

progeny virus, hence this protein is essential for viral replication (Sodroski et al., 1986). The N-terminal 47 amino acids comprise the activation domain, which contains a cysteine-rich region and a hydrophobic core segment. The C-terminal (basic) domain is involved in TAR binding, and serves as a nuclear localisation signal (Siomi et al., 1990; Dingwall et al., 1989). In addition to transactivation, Tat has a number of other effects including induction of apoptosis in uninfected bystander cells (Li et al., 1995).

Rev

Rev (regulator of expression of viral proteins) is a 116 amino acid 19 KDa protein encoded by two exons. Rev is expressed during the early phase of infection, and is localised primarily in the nucleus but shuttles rapidly between the nucleus and cytoplasm (Zhang et al., 1996). Its function is to bind to partially spliced and unspliced viral RNA transcripts that accumulate in the nucleus, and export them through the nuclear pore complex into the cytoplasm (Fritz et al., 1995; Alonso and Peterlin, 1999). Binding of Rev to viral RNAs is dependent upon the presence of a 250 nucleotide RNA secondary structure named the Rev response element (RRE), which is located within *env* and is present in all partially spliced and unspliced viral RNA transcripts (Malim et al., 1989). Rev contains two functional domains, an arginine-rich domain, and an activation domain. The N-terminal arginine-rich domain mediates nuclear localisation and RNA binding, and is flanked by sequences that facilitate multimerisation (Hammerschmid et al., 1994). The activation domain is leucine-rich and contains a nuclear export sequence (Fischer et al., 1995).

1.2.4 Accessory Proteins

Nef

Nef (negative factor) is a 206 amino acid, 27 KDa membrane-associated phosphoprotein. The *nef* gene is unique to primate lentiviruses, and overlaps with the 3' LTR. Nef serves a number of functions in infected cells, although the best characterised is the downregulation of CD4 and major histocompatibility complex (MHC) class I molecules from the cell surface (Rhee and Marsh, 1994; Mangasarian and Trono, 1997; Garcia and Miller, 1991). After translation, Nef is myristoylated near the N-terminus and directed to the inner surface of the plasma membrane. Nef then attaches to the cytoplasmic tail of CD4, and associates CD4 with adapter protein complexes in clathrin-coated pits, leading to CD4 internalisation (Aiken et al., 1994; Greenberg et al., 1997). Following endocytosis Nef interacts with β -COP in the endosome, thereby targeting CD4 for lysosomal degradation (Benichou et al., 1994; Piguet et al., 1999). This downregulation of CD4 reduces the formation of intracellular complexes between CD4 and gp120. Downregulation of MHC class I molecules provides some protection for virally infected cells against recognition by cytotoxic T lymphocytes (Collins et al., 1998). Other functions of Nef include the enhancement of virion infectivity (Chowers et al., 1995; Chowers et al., 1994), and modulation of cellular activation pathways (Swingler et al., 1999). Small amounts of Nef are also incorporated into virions (Pandori et al., 1996). Viral isolates from a few individuals who have been infected with HIV-1 for many years without progressing to AIDS have inactivating mutations in *nef* (Rhodes et al., 2000; Deacon et al., 1995). Thus in some cases Nef may be important for disease progression.

Vif

The *vif* (viral infectivity factor) gene overlaps with both the *pol* and *vpr* genes in the centre of the viral genome. All lentiviruses with the exception of EIAV encode a Vif accessory protein (Kawakami et al., 1987). Vif is a 192 amino acid, 23 KDa cytoplasmic protein, present within infected cells in both membrane associated and cytosolic forms. The primary function of Vif is to enhance the infectivity of viral particles. This augmentation is highly dependent upon producer cell type, and independent of target cell type (Gabuzda et al., 1992; Sodroski et al, 1986). When Vif is not expressed then virions produced by “non-permissive” cells such as primary lymphocytes and macrophages have altered core structures, and fail to reverse transcribe correctly in new host cells (Vonschwedler et al., 1993; Borman et al., 1995). Vif overcomes the antiviral action of CEM15 (also known as apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G: APOBEC3G), a cytidine deaminase only present in non-permissive cell types (Sheehy et al., 2002; Zhang et al., 2003). CEM15 is incorporated into virions and induces hypermutation during reverse transcription of the viral genome in the newly infected host cell (Sheehy et al, 2002; Harris et al., 2003; Lecossier et al., 2003). Therefore CEM15 may represent part of an innate antiretroviral immune effector mechanism that is inactivated by Vif. Vif may also play a role in virion assembly and maturation, possibly by affecting Gag processing (Simm et al., 1995; Borman et al, 1995), and it is incorporated into virions.

Vpr

The *vpr* (viral protein R) gene encodes a 96 amino acid, 14 KDa protein that is incorporated at high levels into virions (Cohen et al., 1990). Virion incorporation of

Vpr is mediated by an interaction with the C-terminus of p6 within Pr55^{gag}. A number of different functions have been proposed for Vpr. These include aiding transport of the preintegration complex into the nucleus (Popov et al., 1998a; Popov et al., 1998b), arresting infected cells in G2 of the cell cycle (Poon et al., 1998), and stimulation of gene expression driven by the HIV LTR (Cohen et al., 1990). Vpr has two helical domains that have been implicated in dimerisation, one near the N-terminus and the other near the centre of the protein (Schuler et al., 1999; Zhao et al., 1994). The C-terminus of Vpr contains an arginine-rich domain. Vpr forms part of the preintegration complex, and is particularly important in facilitating replication within terminally differentiated cells such as monocyte-derived macrophages (Balliet et al., 1994; Heinzinger et al., 1994). The significance of Vpr arresting cells in G2 is not clear, but this may represent a mechanism for prolonging the enhanced LTR-driven transcription that is associated with this stage of the cell cycle (Felzien et al., 1998). Vpr has also been shown to have a wide range of other properties such as the formation of ion channels (Piller et al., 1998), instigation or suppression of apoptosis (Ayyavoo et al., 1997; Stewart et al., 1997), and reduction of mutation rates during reverse transcription (Mansky et al., 2000; Mansky, 1996).

HIV-2 carries a *vpx* gene that is homologous to *vpr*, which may have been obtained by acquisition of the SIV_{agm} (African green monkey) *vpr* gene by recombination (Sharp et al., 1996; Tristem et al., 1998).

Vpu

vpu (viral protein U) is unique to HIV-1/SIV_{cpz}, and no similar gene is present in HIV-2 or other SIVs (Cohen et al., 1988; Huet et al., 1990; Strebel et al., 1989). Vpu is a multimeric, 81 amino acid, 16 KDa integral membrane phosphoprotein that is

translated from *vpu-env* bicistronic RNA (Schwartz et al., 1990). The N-terminus comprises a hydrophobic membrane-spanning domain, and the C-terminus consists of an approximately 50 amino acid hydrophilic cytoplasmic domain (Maldarelli et al., 1993; Strebel et al., 1989). Serine residues at amino acid positions 52 and 56 are phosphorylated by casein kinase II (Schubert et al., 1994). Vpu has been associated with two major functions, degradation of CD4 (Chen et al., 1993; Lenburg and Landau, 1993; Vincent et al., 1993), and enhancement of virion release from the plasma membrane (Geraghty and Panganiban, 1993). Vpu mediated degradation of CD4 is dependent upon binding to the cytoplasmic tail of CD4. gp160 is thereby disassociated from CD4 and can be transported to the cell surface for incorporation into virions (Chen et al., 1993). The mechanism by which Vpu enhances virion release is less well characterised. However, it appears to be independent of both Env and CD4 (Geraghty and Panganiban, 1993), and may involve the ability of Vpu to form cation-selective ion channels (Ewart et al., 1996; Schubert et al., 1996).

1.3 The HIV-1 life cycle

The first step in the HIV-1 replicative cycle is virion adsorption to a target cell expressing CD4 and a coreceptor molecule (Fig. 1.5). This is followed by fusion of the viral and host cell membranes, leading to internalisation and uncoating of subviral particles, and exposure of the viral nucleoprotein complex.

Reverse transcription of the viral genome occurs within this complex soon after entry into the cell, and results in the production of linear, double stranded viral DNA. The preintegration complex is then transported from the cytoplasm into the nucleus and the viral genome is integrated into the host cell chromosome. Transcription of the

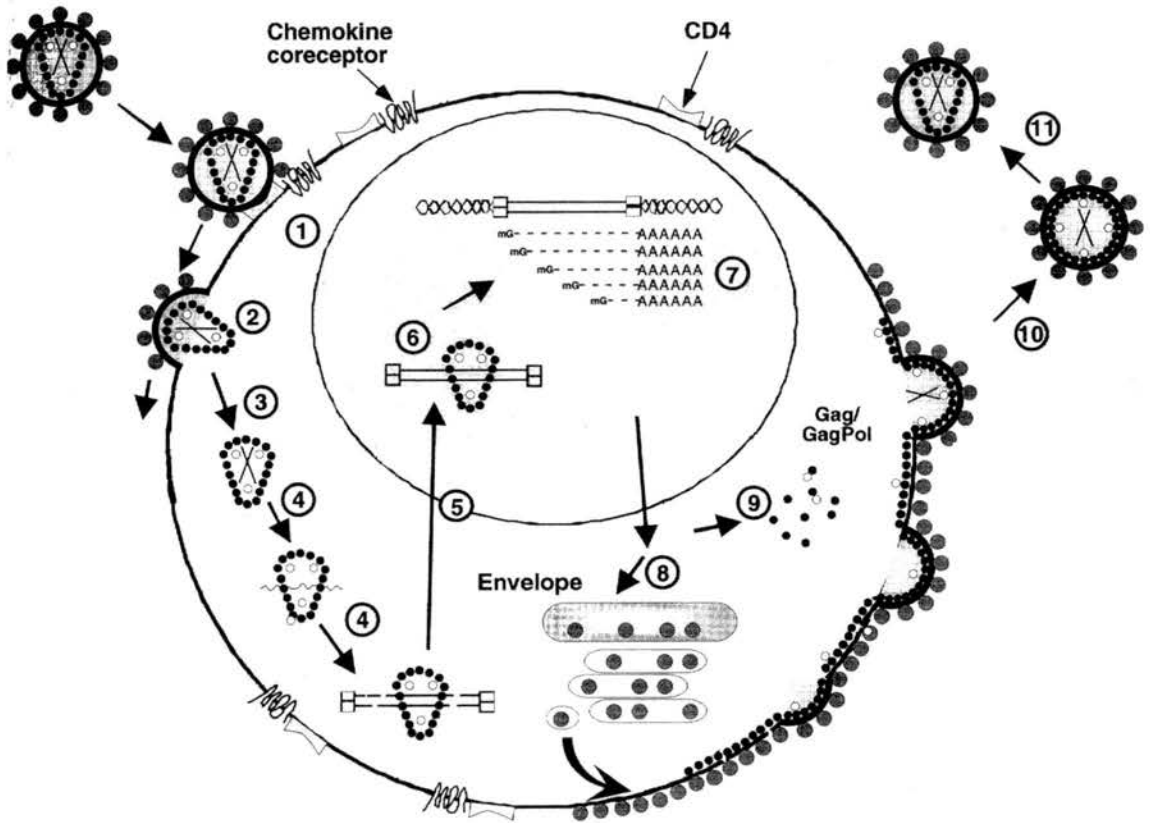
integrated provirus and subsequent modifications by host cell enzymes results in the production of both full-length and spliced viral transcripts, which are transported into the cytoplasm for translation into proteins. At the host cell membrane genomic RNA is packaged into the newly forming virion. The virion then buds from the cell and is released into the extracellular matrix.

Protease mediated cleavage of Gag and Gag-Pol precursor proteins during or soon after budding results in core condensation, and the new virion is then prepared for infection of a new cell.

Figure 1.5. The HIV-1 Life Cycle.

Early stage: (1) HIV-1 adsorbs to cell and interacts with both CD4 and a coreceptor molecule via gp120. (2) Membrane fusion mediated by gp41 allows entry of the subviral particle into host cell. (3) Partial uncoating takes place during or soon after entry. (4) Reverse transcription occurs in the cell cytoplasm. (5) The linear, double stranded DNA product is transported into the nucleus within a preintegration complex. (6) The integrase enzyme (open circles) mediates integration of viral DNA into the host cell chromosome.

Late stage: (7) Integrated viral DNA serves as a template for transcription of viral mRNAs, which are transported into the cytoplasm and translated into proteins. (8) and (9) Envelope and Pr55^{Gag}/Pr160^{Gag-Pol} are transported to the plasma membrane via independent pathways. (10) Full-length viral genomic RNA and other virion components are also assembled at the cellular membrane. The progeny virus then buds from the cell as an immature particle. (11) Proteolysis of virion polyproteins by virally encoded protease leads to the production of mature infectious particles. Figure modified from (Goff, 2001).



1.3.1 Adsorption and fusion

In contrast to enveloped viruses such as influenza that are internalised prior to membrane fusion, HIV-1 fusion can occur at the cell surface and is not dependent upon pH changes associated with endocytosis. A large number of factors can lead to adsorption of HIV-1 virions to host cells, but the first requisite interaction for cellular entry of naturally occurring HIV-1 isolates is between gp120 and CD4 (Maddon et al., 1986; Dalgleish et al., 1984). CD4 has four extracellular immunoglobulin-like domains named D1 to D4 (D4 being proximal to the cellular membrane). Residues in D1, corresponding to the complementarity-determining region 2 of an immunoglobulin light chain variable domain, form a charged ridge and are responsible for binding gp120 (Kwong et al., 1998). In particular, a phenylalanine side chain at position 43 in CD4 inserts into a cavity in gp120, making multiple contacts with discontinuous residues in gp120 (Kwong et al., 1998). The determinants in Env for CD4 binding map mostly to regions of C3 and C4, but also to a discontinuous and conformation dependent domain of gp120. Binding of CD4 induces conformational changes in the core of gp120 (Myszka et al., 2000), and exposes/forms the coreceptor binding site. Each gp120 monomer in the trimeric envelope glycoprotein spike contains a binding site for CD4, and engagement of a single monomer by one CD4 molecule is sufficient to induce conformational changes in all three gp120 monomers in the spike (Salzwedel and Berger, 2000). Other changes in Env at this point may include rearrangements of the N- and C- helices of gp41 (exposure) necessary for membrane fusion (Furuta et al., 1998; Sattentau and Moore, 1991). CD4 has flexible regions between D2 and D3, and also between D4 and the cellular membrane, which are important for its function as a lentiviral

receptor (Yachou and Sekaly, 1999; Healey et al., 1990; Moir et al., 1996). It is possible that bending of CD4 at these points allows gp120 to “approach” CD4 laterally for binding, and also facilitates the subsequent contacts between gp120 and the coreceptor molecule (Clapham and McKnight, 2002).

Engagement of a second receptor or “coreceptor” molecule is also necessary for entry of HIV into cells. HIV coreceptors are all members of the 7-transmembrane G protein coupled receptor superfamily, and the two most frequently used by HIV-1 isolates are CCR5 and CXCR4 (Feng et al., 1996; Alkhatib et al., 1996b; Choe et al., 1996; Deng et al., 1996; Doranz et al., 1996; Dragic et al., 1996). Different HIV-1 isolates vary greatly in the specifics of their interactions with coreceptor molecules, but generally the N-terminus of CCR5 and the second extracellular loop of CXCR4 contain the most important determinants for binding of gp120 (Hill et al., 1998; Lu et al., 1997). Differential coreceptor usage is an important determinant of HIV-1 cellular tropism. Therefore the structure and cellular distribution of different coreceptors, as well as their role in HIV entry and tropism will be discussed later in more detail (section 1.7.2)

The conformational changes in both gp120 and gp41 associated with CD4 and coreceptor binding result in exposure of the fusion peptide at the N-terminus of gp41. This fusion peptide inserts into the cellular membrane, making gp41 an integral protein in the two membranes. Further interactions within the gp41 trimer result in the formation of a thermodynamically stable six helical bundle (Chan et al, 1997). Energy released during this transition is thought to drive fusion of the viral and host cell lipid bilayers. Several trimeric envelope spikes are involved in the induction of a

fusion pore, but the exact number required is not known (Clapham and McKnight, 2002). In influenza virus three to six HA trimers are necessary for this process (Danieli et al., 1996). HIV-1 envelope proteins present on the surface of infected cells can also mediate fusion directly with adjacent cells if they express the appropriate receptors, resulting in the formation of syncytia.

1.3.2 Reverse transcription

Subsequent to membrane fusion the viral core enters the cell cytoplasm. Events immediately after membrane fusion and prior to reverse transcription (collectively referred to as uncoating) are poorly defined but involve partial disruption of the viral core to form a preintegration complex (PIC). The PIC is composed of MA, NC, Vpr, RT, IN, and viral RNA. This nucleoprotein complex is involved in reverse transcription, nuclear import, and possibly integration of the double stranded DNA viral genome (Bowerman et al., 1989; Bukrinsky et al., 1993b; Heinzinger et al., 1994; Freed and Martin, 1994).

Retroviral reverse transcription is performed by the RT enzyme, and occurs in a number of discrete steps (Gilboa et al., 1979). The primer used by HIV-1 is tRNA^{Lys3}, which is carried within the virion (Jiang et al., 1993). Primer annealing to the primer binding site (pbs) located near the 5' end of the viral genome is facilitated by NC (Rodriguezrodriguez et al., 1995; Huang et al., 1997). The primer is elongated towards the 5' end of the genomic RNA to synthesise U5 and R sequences, and thereby generate minus strand strong-stop DNA. RNase H activity of RT then degrades RNA outwith the pbs in the resultant DNA/RNA heteroduplex. This is followed by a strand transfer or "jump" of the minus strand strong-stop DNA from

the 5' to the 3' end of the viral genome, guided by complementary R sequences. Elongation of minus strand strong-stop DNA then continues to the 5' end of the genome, and as polymerisation proceeds RNase H activity degrades the entire RNA genome, with the exception of short purine-rich sequences that serve as primers for plus strand DNA synthesis. In HIV there are two such regions, the standard polypurine tract (PPT), which is located just 5' to U3 in the 3' LTR and is common to all retroviruses, and an additional central PPT (cPPT) located near the middle of the genome. The cPPT is required for optimal HIV reverse transcription (Charneau et al., 1992). Plus strand synthesis is initiated at the PPTs using minus strand DNA as template. The original tRNA^{Lys3} primer and RNA PPTs are degraded, and transfer of plus strand strong-stop DNA occurs, facilitated by the complementary pbs sequences at each end of the genome. Elongation of both plus and minus strands then proceeds to completion. Plus strand synthesis initiated at the cPPT terminates at the end of the minus strand. However, the transferred plus strand strong-stop DNA terminates at a sequence known as the central termination signal (Charneau et al., 1994), which is situated downstream of the cPPT. This causes the displacement of approximately 99 nucleotides, and the resultant DNA "flap" has been implicated as a determinant of nuclear import in terminally differentiated cells (Zennou et al., 2001; Zennou et al., 2000). Any discontinuities in the final double stranded viral DNA are likely resolved post integration (Miller et al., 1995).

1.3.3 Integration

After reverse transcription the double stranded DNA is transported into the nucleus within the preintegration complex. Unlike simple retroviruses, lentiviruses do not require the breakdown of the nuclear membrane associated with mitosis to gain

access to cellular DNA. Therefore lentiviruses can productively infect nondividing and terminally differentiated cells such as macrophages and microglia. The specific process by which the HIV-1 preintegration complex is transported across the intact nuclear membrane is not known, but *cis*-acting components of the double stranded viral DNA and nuclear localisation signals present in several proteins within the PIC have been implicated in this process (Zennou et al, 2001; Zennou et al, 2000; Gallay et al, 1997; Popov et al, 1998a; Popov et al, 1998b).

The steps required for integration of HIV into the host cell DNA are performed by IN. Firstly 3' end processing takes place, whereby IN removes the two terminal nucleotides at the 3' ends of the blunt ended LTRs (Roth et al., 1989). This endonucleolytic cleavage occurs at the 3' end of a highly conserved CA dinucleotide, and produces 3' hydroxyl groups. The resultant 3' OH ends are used in a strand transfer reaction to attack the phosphodiester bonds of host cell target DNA (Fujiwara and Mizuuchi, 1988). This forms new phosphodiester bonds between the viral and host DNA, and leaves two staggered nicks in the double stranded host DNA. The result is a gapped intermediate, where several nucleotides immediately flanking the 5' ends of the viral DNA are missing. Host enzymes repair these gaps and the provirus is then both fully integrated, and flanked by 5 bp of duplicated target site sequence.

Retroviral proviruses appear to be inserted into the host genome in a nearly random fashion with little or no sequence specificity. However, slight deviations from a completely random distribution of integration sites have sometimes been noted (King et al., 1985). Investigations using *in vitro* systems have identified some

potential “hot spots” for integration (Kitamura et al., 1992; Shih et al., 1988). Other *in vivo* observations have identified a slight preference for integration into transcriptionally active genes (Scherdin et al., 1990; Mooslehner et al., 1990), and open or accessible chromatin structures (Craigie, 1992; Pryciak et al., 1992). Because some sites in the host genome are transcriptionally silent whereas others are expressed at high levels, the specific site of proviral integration can greatly affect subsequent expression of the viral genome (Feinstein et al., 1982; Akroyd et al., 1987).

1.3.4 Transcription of viral DNA

Events following integration mark the beginning of the “late” stage of retroviral replication. The integrated provirus is expressed in a similar manner to cellular genes, and is dependent upon host cell proteins for transcription. Promoter and enhancer sequences within the 5' LTR recruit RNA polymerase II and other cellular proteins to the transcriptional start site situated at the U3-R border. This results in the production of full-length viral transcripts. Cellular machinery adds a 5' cap then cleaves and polyadenylates the RNA at the R-U5 border of the 3' LTR. Resultant full-length transcripts are either packaged into new virions, used as templates for the translation of Gag and Gag-Pol precursor proteins, or spliced to produce mRNA for translation of Env and auxiliary proteins. Splicing of HIV-1 RNA is complex, and over 30 different HIV-1 transcripts have been identified in productively infected cells (Schwartz et al, 1990; Purcell and Martin, 1993). The promoter clearance and processivity of RNA polymerase II instigated by transcription factors such as NF- κ B and Sp1 is inefficient, and results in only basal levels of transcription. This early transcription leads to expression of Tat, which localises within the nucleus. Tat then

promotes phosphorylation of the C-terminal domain of RNA polymerase II by interacting with both the cyclin T/CDK9 complex, and the TAR stem loop in nascent viral RNAs. The result is a several hundred-fold increase in production of full-length viral transcripts (Alonso and Peterlin, 1999).

HIV-1 gene expression results in the translation of “early” and “late” gene products. Initially, multiply spliced viral transcripts are exported from the nucleus by normal cellular pathways, allowing translation of the early proteins Tat, Rev, and Nef. After translation Rev is imported into the nucleus and facilitates export of full-length and partially spliced viral RNAs into the cytoplasm by interaction with both the RRE and the nuclear export machinery. This allows the translation of the late gene products Pr55^{Gag}, Pr160^{Gag-Pol}, Env, Vpu, Vpr, and Vif. The transition from early to late gene expression is associated with the accumulation of a critical level of Rev (Pomerantz et al., 1990; Kim et al., 1989).

1.3.5 Assembly and Budding

HIV-1 virions are assembled at the plasma membrane. Virion assembly is driven primarily by the Pr55^{gag} precursor protein, and expression of this polyprotein in isolation of the rest of the HIV-1 genome is sufficient to produce non-infectious virus-like particles (Freed and Martin, 2001). It is estimated that approximately 2000 interacting Gag polyproteins are involved in the formation of a single retroviral virion (Stromberg et al., 1974). Pr55^{gag} and Pr160^{gag-pol} assemble and are anchored to the inner side of the plasma membrane by a myristic acid attached to the N-terminus of MA. NC incorporates dimers of genomic RNA into the newly forming virion through an association with the packaging signal. Vpr is included via an interaction

with p6 (Paxton et al, 1993). Incorporation of envelope glycoproteins may be aided by a specific interaction between MA and the cytoplasmic domain of gp41 (Freed and Martin, 1996; Murakami and Freed, 2000). Cellular components such as cyclophilin A and tRNAs, as well as integral host cell membrane proteins, also form part of the budding virion. Cleavage of Pr55^{gag} and Pr160^{gag-pol} by PR then causes the dramatic structural rearrangements associated with virus maturation.

1.4 Pathogenesis of HIV-1

1.4.1 Transmission of HIV-1

Early epidemiological observations suggested that homosexual sex between men was a significant transmission route for HIV. It was subsequently noted that HIV could also be transmitted through heterosexual sex (Harris et al., 1983). In developed countries heterosexual HIV transmission has gradually become more prevalent, and the great majority of newly acquired HIV infections worldwide are via this route.

HIV transmission can also occur from mother to child. Vertical transmission may occur by intrauterine infection of the foetus, through exposure to maternal blood and genital secretions at birth, or postnatally via breast-feeding. In the absence of appropriate treatment and medical advice the transmission rate from mother to child can be as high as 60% in some geographical areas, although rates of around 30% are more common (Duong et al., 1999; Peckham and Gibb, 1995). Where available, the use of antiretroviral therapy for both mothers and neonates, along with caesarean section delivery and advice against breast-feeding has resulted in a drastic decrease in vertical transmission (Duong et al, 1999).

A further mode of HIV transmission is parenteral. This may occur during transfusion with infected blood products, by exposure to contaminated blood through re-use of needles or syringes among intravenous drug users, or in health care facilities where instrument sterilisation is inadequate (Quinn, 1996). A significant number of haemophiliacs and blood transfusion recipients have been infected by receiving either blood or blood products (such as factor VIII and factor XI) contaminated with HIV. The introduction of compulsory screening of donated blood and heating of blood products has resulted in a dramatic decrease in this route of HIV transmission (Lackritz et al., 1995). Moreover provision of sterile injecting equipment has reduced the rates of HIV transmission among intravenous drug users in developed countries (Coates et al., 1996)

The probability of HIV transmission is related to the amount of infectious virus present in the transmitting body fluid. Therefore the transmission risk is greatest when viral loads are high, such as at the peak of viraemia in the acute stage of infection, or later during the symptomatic AIDS phase.

1.4.2 Disease Progression

The natural history of HIV infection can be classified into three phases. The first phase is acute primary infection, characterised by a transiently high viral load. In over 50% of newly infected individuals this is accompanied by a seroconversion illness often described as influenza- or infectious mononucleosis-like symptoms (Schacker et al., 1996). These may include headaches, muscle pains, fever, sore throat, and swollen lymph nodes. Initial symptoms normally last for several weeks but lymphadenopathy, lethargy, and malaise may persist for many months.

Seroconversion can occur any time from several days to several months after primary infection, but antibodies against the viral core or envelope can usually be detected within six to eight weeks (Feinberg, 1996).

Following the acute phase a sharp reduction in viral load ensues, marking the beginning of the “asymptomatic” period of infection. At this point there are few clinical manifestations associated with HIV infection, although some individuals may experience a persistent generalised lymphadenopathy (Feinberg, 1996). The asymptomatic period can last from several months to over ten years in the absence of antiretroviral therapy.

The third stage of infection is the symptomatic phase, or acquired immunodeficiency syndrome (AIDS). This phase is characterised by the presence of immune abnormalities, opportunistic infections, neurological disorders, and immunodeficiency-associated cancers, which ultimately culminate in death of the infected individual.

1.4.3 Viral Dynamics and Immune Responses

Each stage of HIV infection is associated with alterations in viral dynamics accompanied by changes in the anti-HIV immune response (Fig. 1.6). After initial infection there is a burst of viral replication leading to a rapid increase in plasma viraemia, which may reach 10^6 - 10^7 virion copies/ml (Clark et al., 1991; Daar et al., 1991).

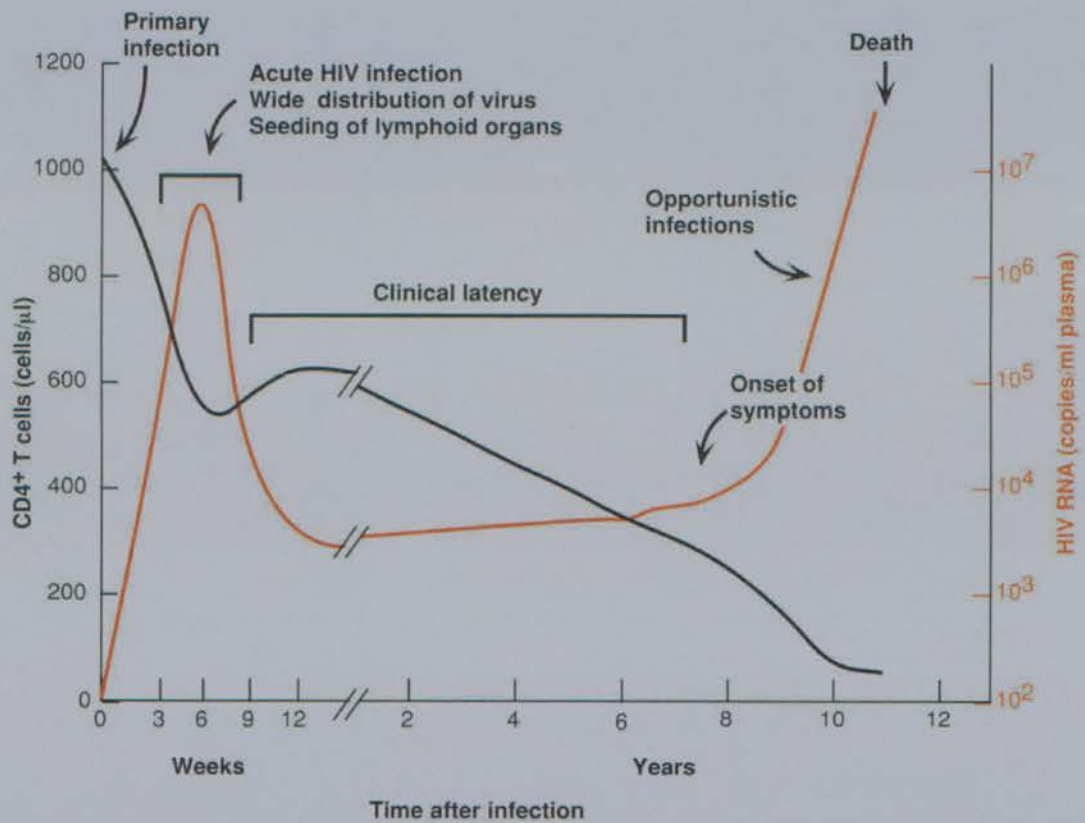


Figure 1.6. Typical Course of HIV Infection. Patterns of CD4⁺ T lymphocyte decline and duration of infection can vary greatly between patients, as do the actual values of viral RNA load. Figure copied from (Fauci and Desrosiers, 1997).

This initial rise in viraemia in the first few weeks of infection is concomitant with a significant decline in CD4⁺ T lymphocytes (Gaines et al., 1990). Specific anti-HIV immune responses can be detected approximately 3-6 weeks after infection (Ho et al., 1985; Gaines et al., 1987).

A reduction in virus titre is associated with the appearance of HIV specific CD8⁺ cytotoxic T lymphocytes (Koup et al., 1994), although factors such as the presence of virus suppressing cytokines and possible exhaustion of suitable CD4⁺ host cells may also contribute to this decline (Phillips, 1996). CD4⁺ T lymphocyte levels then rebound as the asymptomatic stage of infection is entered, but seldom return to preinfection levels.

The asymptomatic period is usually associated with a steady decline in numbers of CD4⁺ T lymphocytes, whereas levels of CD8⁺ lymphocytes remain stable or slightly increase. Viral loads are generally constant throughout this period, but are maintained at these steady state levels by a remarkably dynamic process of viral replication in conjunction with immune-mediated destruction of virions and infected cells (Coffin, 1995; Perelson et al., 1996; Ho et al., 1995; Wei et al., 1995).

Progression to AIDS is associated with a rapid decline in both CD4⁺ and CD8⁺ T lymphocytes, and an increase in viral load. When CD4 counts have dropped from normal levels of approximately 600-1200 cells/ μ l to around 500 cells/ μ l the first symptoms can appear in infected individuals. At CD4 counts of below 200 cells/ μ l the individual is also susceptible to opportunistic infections and neoplasms associated with AIDS (Fauci and Desrosiers, 1997). AIDS is the result of both the progressive depletion of CD4⁺ T lymphocytes which is not matched by a

corresponding increase in the production of new cells, and a number of other factors such as the irreversible destruction of lymphoid tissue (Fauci, 1988; Klatzmann et al., 1984a; Hellerstein, 2002; McCune, 2001; Cohen and Fauci, 2001).

1.4.4 Antiviral Therapy

While the development of a prophylactic vaccine for eliciting protective immune responses against HIV has proven extremely challenging (Nabel, 2001), significant advances have been made in the field of antiretroviral therapy (Richman, 2001). All currently licensed antiretroviral drugs target the RT or PR viral enzymes, and fall into four different categories: nucleoside reverse transcriptase inhibitors (NRTIs), nucleotide reverse transcriptase inhibitors, non-nucleoside reverse transcriptase inhibitors (NNRTIs), and protease inhibitors (PIs). Treatment regimens usually consist of a combination of three drugs, typically comprising 2 NRTIs and a PI or a NNRTI (BHIVA, 2001). These regimens, termed highly active antiretroviral therapy (HAART), usually result in a significant decline in viral load, an increase in CD4 count, and a decrease in mortality (Vella and Palmisano, 2000). However, issues such as expense, bioavailability, toxic side effects, pill burden, and the outgrowth of resistant viral variants make currently available therapies far from ideal. In addition, the majority of HIV infected individuals worldwide have no access to such treatments. A wide range of alternative therapies, designed to target almost every section of the HIV life cycle are currently under development.

1.5 Genetic diversity of HIV-1

One of the most striking features of HIV biology is the remarkable genetic diversity observed between different viral isolates. This high degree of genetic variability was originally observed soon after the discovery of HIV (Hahn et al., 1984; Muesing et al., 1985; Sanchez Pescador et al., 1985; Ratner et al., 1985; Shaw et al., 1984) and it rapidly became clear that intrapatient HIV sequences exhibit evolution over time (Saag et al., 1988; Fisher et al., 1988; Hahn et al., 1986). HIV sequences from a single study subject can be as much as 10% divergent at the nucleotide level. The resulting group of closely related but genetically distinct viral variants has come to be referred to as a "viral quasispecies" (Goodenow et al., 1989). This term is derived from the "molecular quasispecies" (Eigen and Schuster, 1977) originally used to describe a well-defined concept in molecular evolution (Eigen, 1971) whereby selection pressures act upon on a distribution of molecules at equilibrium comprising a master copy surrounded by related sequences. Thus, a viral quasispecies does not conform to this original mathematical definition of a molecular quasispecies (Eigen, 1996; Domingo and Holland, 1997; Jenkins et al., 2001) and other designations such as "viral swarm" have been suggested (Temin, 1989). Despite these inconsistencies a number of parallels can be drawn between the two situations. Therefore "viral quasispecies" or simply "quasispecies" has been loosely adopted in virology to refer to either the related viral variants that have diverged within a single infected individual, or genetically distinct subpopulations of these variants (although it is also used in many other contexts: Smith et al., 1997). This variation may have important implications for evasion of host immune responses, development of resistance to antiviral therapies, and *in vivo* viral adaptation to new cell and tissue types.

1.5.1 Sources of mutation during HIV replication

Genetic variability is determined by the mutation rate per replication cycle, the number of replication cycles within a defined time, and the fixation rate of mutations within a population. HIV-1 replication is complex and involves three enzymes that provide opportunities for mutation to occur. During retroviral replication RT uses viral RNA as a template to produce double stranded DNA, which is then integrated into the host cell genome. RT does not include a 3'→5' exonuclease function that would permit proofreading of misincorporated bases, although several rudimentary proofreading mechanisms have been identified that may partially compensate for this deficiency (Arion et al., 1998; Meyer et al., 1998). The fidelity of HIV RT is therefore low, with estimates ranging from approximately 10^{-4} to 3.4×10^{-5} mutations per base incorporated, corresponding to between approximately 0.34 and 6 errors per genome per round of replication (Roberts et al., 1988; Preston et al., 1988; Ji and Loeb, 1992; Mansky and Temin, 1995). The wide range of these estimates largely reflects differences in experimental designs, with isolated HIV RT having lower fidelity than *in vivo* RT reactions (Mansky and Temin, 1995). The mutation rate of HIV-1 RT is therefore slightly higher but of a similar order to those of RTs isolated from other retroviruses (reviewed in Mansky, 1998). Interestingly, given retroviral genome sizes of approximately 8-12 Kb it appears that these mutation rates are close to the maximum possible while still maintaining genomic and structural integrity within the viral populations.

Investigations into the regularity that different types of error occur during reverse transcription have produced varying estimates depending upon the model system used (reviewed in Preston and Dougherty, 1996). It appears that the overall mutation

rate is similar regardless of whether the template is DNA or RNA, although the specific distribution of mutations is partially dependent upon template molecule type (Preston and Dougherty, 1996). Template position can also have dramatic effects on the fidelity of reverse transcription (Bebenek et al., 1989). Point mutations predominate, with transitions more common than transversions. Several additional types of error are also frequently observed including insertions, deletions, frameshifts, and recombination events (Bebenek et al, 1989; Temin, 1993) (reviewed in Katz and Skalka, 1990; Coffin, 1992; Preston and Dougherty, 1996). Frameshifts are most often represented by the addition or deletion of a single base in a homopolymeric run, and occur with a frequency several fold lower than point mutations (Pathak and Temin, 1990b; Pathak and Temin, 1990a). Genetic rearrangements such as insertions and deletions are less common than frameshifts, and deletions appear to be potentiated by the presence of internal repeats (Delviks and Pathak, 1999). Another well-documented event that can occur during reverse transcription is hypermutation (Vartanian et al., 1991). This is a fairly common phenomenon probably resulting from fluctuations in intracellular deoxyribonucleotide triphosphate ratios where (most commonly) a large percentage of Guanines in the genome are replaced by Adenines (Vartanian et al, 1991; Kim et al., 1996; Martinez et al., 1994). Hypermutation has also been identified in other lentiviruses (Wainhobson et al., 1995) and in hepatitis B virus (Gunther et al., 1997). Finally, recombination can occur between the two copackaged genomic RNA molecules present in the virion, sometimes leading to extremely dramatic sequence changes (see section 1.6). The overall fidelity of reverse transcription can be affected

by a number of additional factors, for example by the presence of some antiretroviral drugs (Julias et al., 1997; Mansky et al., 2002; Mansky and Bernard, 2000).

If the host cell undergoes mitosis after proviral DNA integration then the provirus is replicated in the same manner as other cellular genes. This process is guided by cellular DNA polymerases α and δ (Waga and Stillman, 1998), which include proofreading functions and have a misincorporation rate of less than 1 for every billion bases copied. Therefore mutations incorporated during cell division do not significantly contribute to the high level of *in vivo* HIV genetic variation. However, it is possible that very occasionally mutations incorporated during mitosis (especially in rapidly dividing cells) may produce new viral variants or cause reversion of a defective provirus.

New HIV genomes and mRNAs are produced within the cell by RNA polymerase II, which lacks 3'→5' exonucleolytic proofreading activity and further contributes to mutations incorporated during HIV replication. The specific mutation rate for this enzyme is currently unknown but is likely to be of a similar order to that of prokaryotic RNA polymerases i.e. one error in every 10^4 - 10^5 nucleotides polymerised (Erie et al., 1993; Blank et al., 1986; Libby and Gallant, 1991). Hence misincorporations by cellular RNA polymerase II may be as important for generation of HIV diversity as RT induced errors.

In addition to the polymerase errors outlined above, mutation of retroviral DNA or RNA may also result from other factors. These include spontaneous chemical "decay" of RNA or DNA (Lindahl, 1993) leading to nucleotide misincorporations

during double stranded viral DNA synthesis, and aberrant RNA editing (Dougherty and Temin, 1986).

It has been suggested that certain areas of the HIV genome have especially high mutation rates and that the presence of these mutational "hotspots" may explain why specific regions, such as within *env*, are hypervariable as compared with the rest of the HIV genome (Bebenek et al, 1989). Other studies have failed to find these inherent differences in mutation rate (Ji and Loeb, 1994), and instead highlighted the importance of factors such as selection pressures in the fixation of mutations and development of variation in different subgenomic regions.

1.5.2 Viral turnover in HIV infected individuals

A high rate of replication also contributes to rapid evolutionary changes. HIV establishes a persistent infection in humans, which often lasts for over 10 years before culminating in AIDS. It has long been known that HIV-1 actively replicates to some extent throughout the course of infection (Embretson et al., 1993; Pantaleo et al., 1993). However the magnitude of this replication was not fully appreciated until the development of potent antiretroviral therapies and new techniques for the measurement of viral RNA, such as branch chain DNA and RT-polymerase chain reaction assays (RT-PCR). Use of a protease inhibitor or a non-nucleoside reverse transcriptase inhibitor almost completely suppressed viral replication in treated patients. Careful measurement of viral loads in these patients after drug administration therefore allowed the rates of clearance for free virions in the peripheral blood and productively infected cells to be calculated. Results indicated that the half-life of HIV virions in plasma is approximately 6 hours or less, and that

of productively infected CD4⁺ cells contributing to this virus is approximately 1.5 days (Ho et al, 1995; Wei et al, 1995; Perelson et al, 1996; Perelson et al., 1997). As there is little natural short-term variability in plasma HIV RNA levels over time (Hughes et al., 1997c) these figures for the clearance rate of virions and infected cells must be equivalent to the numbers produced. Therefore approximately 10^8 - 10^9 free virions and 10^6 - 10^7 productively infected cells must be produced and cleared each day solely to generate the virus observed in plasma. In the body as a whole the corresponding figures are likely to be several times higher. Thus the clinically latent period of HIV-1 infection is not one of virological latency but rather a dynamic equilibrium characterised by very high levels of viral replication offset by an equally prodigious immune response.

These interpretations are consistent with data from patients undergoing treatment with antiviral drugs that require only a small number of mutations for the development of high level resistance. For example even single codon changes at specific positions (such as position 181) in the reverse transcriptase gene can decrease the susceptibility of HIV to the non-nucleoside reverse transcriptase inhibitor nevirapine by over 100-fold (Richman et al., 1994). Use of nevirapine was found to cause a rapid drop in HIV RNA levels within the first 1-2 weeks of therapy, but after this period viral RNA levels rebounded close to pretherapy values (Wei et al, 1995). The new viral variant that emerged during this period was highly resistant to nevirapine and by 4 weeks after the initiation of therapy the plasma HIV RNA consisted of an estimated 100% mutant virus (Wei et al, 1995). Virus that appeared during this time must either have been pre-existing in the HIV population within the patient before therapy, or have rapidly evolved after initiation of drug treatment.

Investigations such as this represent independent confirmation of the extremely rapid plasma virus turnover, and demonstrate the rate at which previously minor viral variants can emerge and predominate within the plasma of infected individuals. Further data from this study came from the investigation of proviral DNA in PBMCs. In contrast to plasma HIV RNA the accumulation of drug resistant mutant virus in PBMC proviral DNA was a prolonged process, leading to half-life estimates for these PBMCs of 50-100 days (Wei et al, 1995). This is compatible with estimates based on quantification of total HIV proviral levels in patients starting potent antiretroviral therapy (Stellbrink et al., 1996; Perelson et al, 1997). It therefore appears that the majority of HIV virions in plasma are produced by a relatively small number of HIV DNA-positive cells. One explanation for this finding is that most HIV DNA in PBMCs is defective, and does not significantly affect the lifespan of the cell or contribute to plasma virus.

An obvious implication of this consistently high rate of viral turnover is that, when combined with the error prone nature of viral replication, new genetic variants are frequently generated. Therefore it is estimated that every possible single base mutation and a large proportion of possible double mutations are produced every day in untreated infected individuals (Coffin, 1995).

1.5.3 Fixation rate of mutations in HIV

In addition to the frequency that new genetic variants are generated, the rate at which specific mutations are fixed within a population is also a major factor in determining the overall rate of evolutionary change. Clearly, genetic mutations can be deleterious, neutral, or advantageous to an organism. In coding sequences non-synonymous

mutations (those that alter the primary amino acid sequence) can compromise structural or functional constraints manifest in the encoded proteins. Synonymous changes can also be detrimental in certain circumstances. For example they may alter RNA secondary structures that are essential for normal functions or involved in replicative processes. These types of mutation can lead to organisms that are non-replication competent, or have a selective disadvantage as compared to the parental phenotype. Such debilitated variants contribute relatively few or no progeny to future generations and are therefore numerically reduced or removed from the gene pool. This process is termed negative (or purifying) selection. Mutations may also be neutral, and have no effect on the organisms chances of survival and persistence within the population. Selectively neutral mutations can still become fixed by the random process of neutral genetic drift, or by genetic linkage with advantageous mutations. Other genetic alterations may endow some form of selective advantage and lead to fitness gains. Variants with these beneficial characteristics tend to contribute more progeny to future generations and are therefore subject to positive selection.

Different genes vary widely in their tolerance to genetic changes (Page and Holmes, 1998b). The extensive diversity within global HIV-1 isolates is testimony to the ability of HIV to endure mutations at numerous nucleotide positions without a significant loss of fitness. The largest differences in major open reading frames are found within the *env* gene, followed by *gag* and then *pol*. These differences are the product of varying selection pressures acting upon different genes, and are also reflected in viral isolates within single infected individuals (Rodrigo et al., 2001).

There has been some debate over the relative contributions of random processes (genetic drift) and selection pressures in driving the fixation of mutations during HIV infection (Rouzine and Coffin, 1999; Brown, 1997; Brown and Richman, 1997). These are often referred to as stochastic and deterministic models respectively. The fraction of a viral population within an individual that contributes to the gene pool of the next generation of virus is referred to as the effective population size (N_e), as opposed to the total population size (N). Behaviour of a quasispecies (or indeed any genetically diverse population) is highly dependent on the value of N_e (reviewed in Domingo and Holland, 1997; Rouzine et al., 2001). In small effective populations the fate of viral variants, regardless of their relative fitness, can be heavily influenced by random events (stochastic behaviour). By contrast, in large effective populations undergoing many rounds of replication, random events have less impact and the ultimate fate of the population is instead governed primarily by the relative fitness of its component variants (deterministic behaviour).

Genetic bottlenecks can occur at various points of HIV infection, such as during transmission or drug therapy. These contractions in total population size may increase the importance of stochastic events. Reduction in N_e during the steady state of infection can also take place and lead to similar stochastic behaviour. This may be the result of relatively few productively infected cells contributing to the next generation of virus, despite the usually large total population sizes at this point. Alternatively, frequent genetic bottleneck events in viable virus, or compartmentalisation leading to subdivision of the total viral population into independent lineages may also lead to reductions in N_e . Evidence from *env* sequences has lent weight to the possibility that effective population sizes of HIV in



infected individuals may be several orders of magnitude lower than the total number of infected cells (Brown and Richman, 1997; Brown, 1997). The extent of a particular viral variant's selective advantage also has an effect on its prevalence in future generations. Neutral or nearly neutral mutations tend to be governed by stochastic effects, whereas those conferring large selective advantages or disadvantages behave more deterministically (Rouzine and Coffin, 1999). It may therefore be expected that under an extreme selection pressure such as HAART, stochastic events would be greatly outweighed by the large selective advantage conferred by mutations associated with antiviral resistance.

The large number of new viral variants produced throughout HIV infection supply extensive raw material for the subsequent actions of natural selection. Perhaps the clearest example of selective forces influencing the prevalence of HIV variants *in vivo* is provided by the introduction of powerful inhibitors of viral replication from external sources, in the form of antiviral drugs. There are numerous cases where treatment with antiretroviral agents has led to the rapid outgrowth of resistant virus, which is by definition more fit than the drug susceptible variants that generally predominate up to this point (Wei et al, 1995; Schuurman et al., 1995; Loveday et al., 1995; Ives et al., 1997). Immune responses also drive viral evolution by selecting for variants with reduced sensitivity to neutralising antibodies and cytotoxic T cells (CTLs). Antibody neutralisation escape mutants have been identified during HIV-1 infection (Arendrup et al., 1992; Lathey et al., 1997). Likewise CTL escape mutants have been described (Phillips et al., 1991; Borrow et al., 1997; Goulder et al., 2001; Price et al., 1997), and in some cases emerge quickly after primary infection (Borrow et al, 1997; Price et al, 1997).

Comparing viral heterogeneity in study subjects with qualitatively different anti-HIV immune responses has also provided indirect evidence for the impact that immune pressures can have on viral diversity. Individuals infected with HIV typically progress to AIDS within ten years. However, some individuals progress much more quickly (rapid progressors), while others remain asymptomatic for more than ten years (slow progressors, or long term non-progressors). While many viral and host factors can influence the rate of disease progression, some studies have identified stronger CTL responses to HIV-1, and higher neutralising antibody titres in slow progressors than in individuals with a typical rate of progression (Pantaleo et al., 1995; Cao et al., 1995; Betts et al., 1999; Montefiori et al., 1996). Rapid progressors often lack these effective immune responses and have weak or no neutralising antibodies to HIV-1 (Liu et al., 1997). Differences in the rate of disease progression also correlate with viral diversity. Slower progressors harbour virus with greater genetic diversity than that found in rapid or typical progressors (Wolinsky et al., 1996; Delwart et al., 1997; Liu et al., 1997). Higher ratios of non-synonymous to synonymous site mutations have also been identified in these individuals (Lukashov et al., 1995), indicating that the prevailing selection pressures differ between the groups of study subjects.

Taken together these results suggest that immune responses are an important factor in the fixation rate of new mutations. When effective immune responses are sparse or absent, viral diversity is much lower than is the case when vigorous immune responses cause frequent selective sweeps leading to the outgrowth of previously rare genetic mutants.

Adaptation to infect new cell or tissue types is another potentially important way that selection may act upon specific subsets of HIV-1 variants. Isolation of minor HIV-1 variants during *in vitro* culture in PBMCs has been documented (Vartanian et al, 1991; Meyerhans et al., 1989). Substantial natural variation in viral replicative abilities in T cell lines or macrophages (Schwartz et al., 1989; Fenyo et al., 1988), and in the use of particular chemokine receptors to enter cells in conjunction with CD4 (see section 1.7) is also common between different HIV-1 strains. Therefore the great diversity of HIV-1 variants *in vivo* may allow the virus to adapt to new niches in the form of previously uninfected cell or tissue types.

1.6 Recombination in HIV-1

Recombination involves the generation of a nucleic acid molecule that contains components of sequence from different sources. Some level of genetic exchange between individuals in a population appears to be advantageous to the long-term evolution of most organisms (Maynard Smith, 1978). Many eukaryotes have therefore evolved specialized systems (such as sexual reproduction and recombination) for reshuffling of genetic information. These can include the obligatory contribution of half the progenies required chromosome complement by each parent, and specific mechanisms for the generation of recombinant molecules during meiosis (chiasmata or crossing over). Horizontal gene transfer is also common in prokaryotes, and allows the acquisition of genetic material from contemporaneous members of a population, rather than exclusively from ancestral organisms. Recombination can increase genetic diversity by bringing together new genomic sequence combinations from different origins. Conversely, it can also act as

a repair mechanism, allowing the rescue of defective or debilitated parental genomes by “patch repair”, whereby a defective segment of one parental sequence is provided by its homologous counterpart in the second parental genome.

Retroviral recombination was originally identified in avian retroviruses (Kawai and Hanafusa, 1972; Vogt, 1971), and is now known to be a common feature of retroviral replication in general (Clavel et al., 1989; Wong and McCarter, 1973). The development of cell culture based systems in which replication of vector virus is restricted to a single cycle has provided many insights into the process of retroviral recombination (Anderson et al., 1998; Hu et al., 1997; Hu and Temin, 1990b; Zhuang et al., 2002; Jetzt et al., 2000). A major factor in this high rate of recombination is that retroviral particles contain two molecules of single stranded genomic RNA. Therefore in contrast to members of all other virus families, retroviruses are diploid. The first step in the production of a new retroviral recombinant is co-infection of a cell by two genetically distinct viral variants. One genomic RNA molecule derived from each of the two resulting proviruses must then be copackaged into a single viral particle, followed by infection of a new cell by this heterodimeric virion (Hu and Temin, 1990a). During reverse transcription in the new host cell RT can switch between the two co-packaged RNA templates, resulting in a new provirus with segments of sequence derived from both original variants.

1.6.1 Molecular mechanisms of retroviral recombination

Reverse transcription takes place in the nucleocapsid core particle released into the host cell during virus entry. An integral part of retroviral reverse transcription is two consecutive strand transfer reactions (or "jumps"), which occur during minus- and

plus strand synthesis respectively (Gilboa et al, 1979). Therefore perhaps the most conceptually simple model of retroviral recombination is for the strong-stop DNA to switch template molecules during one or both of these strand transfers. A single molecule of viral RNA can be used as a template for the synthesis of complete proviral DNA via intramolecular strand jumps (Jones et al., 1994). However, minus strand strong-stop DNA can also transfer intermolecularly (between the two copackaged RNA molecules) leading to a provirus with recombinant LTRs (Hu and Temin, 1990b; Panganiban and Fiore, 1988; Vanwamel and Berkhout, 1998). In contrast, plus strand strong-stop DNA appears to transfer only intramolecularly (Yu et al., 1998; Hu and Temin, 1990b). Yet these obligatory strand transfers do not explain recombination within internal regions of the viral genome. Hence two basic models for the mechanism of retroviral recombination have been proposed: The "strand displacement assimilation model" (plus strand recombination) (Junghans et al., 1982), and the "copy-choice model" (minus strand recombination) (Coffin, 1979).

In the strand displacement assimilation model two minus strand DNA molecules are synthesised from one virion, each using a different genomic RNA strand as template. During plus strand DNA synthesis one of the internally initiated strands is displaced by another plus strand transcript. The displaced strand then anneals to the complementary region of the minus strand DNA derived from the second genomic RNA. The intermediate is then resolved, possibly by mismatch repair. This model was originally proposed after the observation of "H"-shaped reverse transcription intermediates by electron microscopy (Junghans et al, 1982) although subsequent evidence for its occurrence has been limited. Thus, while plus strand recombination

may be important in some cases the current consensus is that retroviral recombination is primarily the result of nascent minus strand DNA transfer between RNA templates.

The original "forced copy choice" model for retroviral recombination during minus strand synthesis proposed that breakage of the viral RNA template necessitates switching of nascent DNA to the second undamaged RNA template molecule. Damaging genomic RNA by γ -irradiation of virus particles led to an increased rate of template switching during minus strand DNA synthesis, providing direct experimental evidence for this model (Hu and Temin, 1992). However a number of other *cis*-acting factors also appear to influence strand transfer during minus strand DNA synthesis. The original forced copy choice model has therefore been modified to include these alternative factors. Hence the modified "copy choice" model of retroviral recombination now encompasses any recombination during minus strand DNA synthesis.

A high level of sequence similarity between the nascent DNA molecule and acceptor RNA template facilitates strand transfers, as demonstrated by the close sequence identity required in the terminal R regions for efficient transfer of strong-stop minus strand DNA (Luo and Taylor, 1990), and the much higher rates of homologous rather than non-homologous retroviral recombination (Hu and Temin, 1990a; Stuhlmann et al., 1990; Zhang and Temin, 1993). Pausing of RT, perhaps accompanied by enhanced RNase H-dependent RNA degradation at the pause site may also enhance strand switching (Buiser et al., 1993; Destefano et al., 1992a; Klarmann et al., 1993; Wu et al., 1995; Luo and Taylor, 1990). This pausing may be

influenced by sequence content (especially homopolymeric nucleotide runs) (Klarmann et al, 1993), and RNA secondary or tertiary structures (Klarmann et al, 1993; Destefano et al., 1992b; Klarmann et al, 1993; Klasens et al., 1999). An "interactive hairpin" model has also been proposed to explain the results of an investigation into strand transfers around the TAR stem loop, where most recombination events occurred in the loop or stem regions, rather than the putative RT pause site (Kim et al., 1997). In this model the stem loop melts as reverse transcription progresses through it, allowing a new interaction between the donor and acceptor RNA templates to form. Nascent DNA then hybridises to the acceptor RNA during copying of the descending part of the hairpin, and synthesis continues using acceptor RNA as template.

Many studies into the sites and frequencies of strand transfers during reverse transcription have been based upon *in vitro* systems with purified RT and relatively short virus-derived RNA template molecules. The *in vivo* situation is clearly more complex. For example during retroviral replication reverse transcription takes place on a ribonucleoprotein complex, and the nucleocapsid protein is known to both enhance strand transfer (Negroni and Buc, 1999), and destabilise RNA secondary structures (Guo et al., 1997) thereby reducing structure induced pausing of RT (Klasens et al, 1999). Results obtained from *in vitro* reverse transcription experiments using short, naked RNA templates can therefore differ from the much more complex *in vivo* situation of full-length dimeric RNA genomes in the highly ordered environment of the nucleocapsid core. One example of discordant results based upon reverse transcription of synthetic RNA templates compared with those utilising complete virions or infected cells concerns the mutation rate of

recombination. In isolated model systems misincorporations during cDNA synthesis have been shown to induce RT pausing and promote template switching (Diaz and Destefano, 1996; Palaniappan et al., 1996). Moreover, in similar experiments also using purified RT, non-template-directed mutations (misincorporations, insertions or deletions) occurred at the point of strand transfer with a frequency of 30-50% (Patel and Preston, 1994; Wu et al, 1995). In contrast template switching during reverse transcription in partially disrupted virions (Zhang and Temin, 1994) and infected cells (Zhuang et al, 2002) is highly precise. It therefore appears that template switching during minus strand DNA synthesis is not particularly error prone, and homologous recombination in retroviruses increases diversity mainly by shuffling genetic information from the parental genomes rather than facilitating the generation of new mutations.

An interesting finding of investigations based on single-round infection assays is that homologous recombination appears to occur in distinct viral subpopulations exhibiting "high negative interference" (Hu et al, 1997; Hu and Temin, 1990b; Anderson et al, 1998), whereby multiple recombination events are observed with a much higher frequency than would be expected if each event occurs independently. It was originally thought that such interdependence of recombination events might be evidence for a forced copy choice model, as this model predicts that recombination breakpoints should occur in pairs, corresponding to the two ends of the assimilated DNA segment (Hu and Temin, 1990b). However, subsequent experiments where plus strand recombination was prevented by deletion of the primer binding site in one of the copackaged sequences also demonstrated a higher than expected frequency of multiple recombination events (Anderson et al, 1998). Possible explanations for

these observations include the presence of subpopulations of virus that are especially predisposed to recombination, or that once a recombination event has occurred it increases the likelihood of further strand switching (Anderson et al, 1998; Hu et al, 1997).

In the context of HIV-1, single round infection assays have demonstrated that this virus undergoes approximately two to three recombination events per replication cycle, and these events can occur throughout the genome (Zhuang et al, 2002; Jetzt et al, 2000; Yu et al, 1998). Given the small genome size involved this means that HIV replication is the most recombinogenic process observed in any mammalian related system described to date (Zhuang et al, 2002).

1.6.2 Recombination between global HIV-1 isolates

Retroviral recombination requires the super- or co-infection of a single cell by two or more virions. Early laboratory studies appeared to indicate that this was a rare event, because interference based superinfection inhibition takes place during HIV infection in cell culture systems (mediated mainly by direct receptor blocking or Vpu- and Nef-induced downregulation of viral receptors) (Hart and Cloyd, 1990; Little et al., 1994; Taddeo et al., 1993; Benson et al., 1993; Chen et al, 1993). It was therefore expected that *in vivo* opportunities for recombination between distinct viral strains would be similarly limited. However, a number of potentially recombinant strains of HIV-1 had been identified, most notably the MAL isolate (Alizon et al., 1986), whose recombinant origin was suggested even before establishment of the subtype classification system (Li et al., 1988). After the recognition of phylogenetically distinct HIV-1 genetic subtypes (Louwagie et al., 1993; Louwagie et al., 1995; Myers et al.,

1992), and as more complete genome sequences became available, a surprisingly large number of HIV-1 group M strains also appeared to be intersubtype recombinants (Robertson et al., 1995). Some of these strains had established large local epidemics (Gao et al, 1996b; Carr et al., 1996; Carr et al., 1998; Bobkov et al., 1998; Liitsola et al., 1998; Gao et al., 1998; Liitsola et al., 2000; Montavon et al., 1999), and in certain geographical areas recombinants were more prevalent than "pure" subtypes. For example a recombinant strain derived from subtypes A and E (previously designated subtype E) is prevalent in south-east Asia (Gao et al, 1996b; Carr et al, 1996), and an A/G intersubtype recombinant strain is responsible for the majority of HIV infections in some areas of western Africa (Carr et al, 1998). Therefore superinfection of both individual human hosts and single cells within these individuals by diverse viral strains must be common enough to allow the numerous, and in many cases complex, intersubtype recombination events that have been identified in global HIV-1 isolates.

Intersubtype recombinants appear to be present wherever different HIV-1 subtypes circulate in the same populations (reviewed in Peeters and Sharp, 2000; Thomson et al, 2002). Some of these are unique and only represented by single isolates, while others have reached a high prevalence and wide geographic distribution. A number of recombinant genomes have also undergone subsequent recombination events, leading to complex mosaic patterns and genomes composed of four or more different subtypes (Montavon et al., 2002; Montavon et al, 1999; Gao et al, 1998). Clearly, intersubtype recombinants that are responsible for a large number of infections are as epidemiologically important as pure subtypes. To reflect this fact, and the increasing number of identified HIV-1 group M subtypes, a new system of nomenclature for

subtypes and recombinants has been established (Robertson et al, 2000). In this system intersubtype recombinants that are responsible for an adequate number of infections to be considered of significant epidemiological importance (and conform to several other criteria) are termed "circulating recombinant forms" (CRF), and given names that reflect the origin of their component subtypes. For example the A/E recombinant virus, which is prevalent in south east Asia (Gao et al, 1996b) is now termed CRF01_AE (Robertson et al, 2000). HIV-1 genomes containing sequences originating from more than three different subtypes are denoted complex (cpx). Hence the isolate from Cyprus originally designated subtype I (Kostrikis et al., 1995) (in actuality a recombinant composed of subtypes A, G, H, K, and unclassified regions: Gao et al, 1998) is now termed CRF04-cpx (Robertson et al, 2000). It is estimated that approximately 10% of current global HIV-1 group M isolates are intersubtype recombinants.

In addition to being abundantly evident between group M subtypes, recombination has been identified between members of HIV-1 groups M and O (Peeters et al., 1999; Takehisa et al., 1999). This is significant because groups M and O are the result of separate cross-species transmission events into humans from chimpanzees (*Pan troglodytes troglodytes*) (Gao et al, 1999), and the identification of recombination between the two demonstrates that viable recombinants can be formed between highly divergent strains of HIV-1. It also raises the possibility that future cross-species transmissions of SIV into humans may be followed by recombination with current HIV strains leading to the emergence of novel chimeric HIV/SIV genomes. However, recombination between HIV-1 and HIV-2 viral strains has not been found, even though individuals infected by both of these viruses have been reported

(Rayfield et al., 1988; Grez et al., 1994). Perhaps HIV-1 and HIV-2 are too divergent to efficiently recombine; alternatively recombination between the two may result in severely debilitated or unviable progeny.

The biological significance of these numerous recombinant genomes is unclear. Likewise the evolutionary pressures that allow or select for the emergence of particular circulating recombinant forms in different human populations are unknown. This is partly because, while some biological differences between group M subtypes have been documented (SotoRamirez et al., 1996; Tscherning et al., 1998; Abebe et al., 1999; Gao et al, 1996b; Korber et al., 1994b), they are either equivocal or do not readily explain the prevalence of particular subtypes in different human populations. It may therefore be that the current distribution of subtypes around the world is largely the result of founder events rather than inherent biological differences in the viral strains. If this is the case then intersubtype recombinants and circulating recombinant forms may also be broadly equivalent in terms of important biological characteristics such as transmissibility in different human populations.

1.6.3 Intrasubtype recombination

Generally, in order to directly identify a recombinant the sequence of each parental genome must be explicitly known, or belong to different and well-defined sequence groups. HIV-1 group M subtypes have been comprehensively characterised and are quite distinct, exhibiting 25-35% amino acid sequence differences in their Env proteins (Sharp et al., 1994). It is therefore relatively simple to identify an intersubtype recombinant because different genomic regions will have phylogenetic affinities that switch between representatives (or consensus sequences) of different

subtypes. The discrete breakpoints where these affinities cross over represent the specific genome positions that recombination took place. Identifying recombination events between members of the same HIV-1 group M subtype is more problematic, because the parental sequences are not members of such separate and clearly defined phylogenetic groups. Most intrasubtype recombination events that have been documented to date have therefore involved single infected individuals in small, well-characterised transmission networks. For example intrasubtype recombination has been identified in both a transfusion recipient (Diaz et al., 1995) and an intravenous drug user (Wang et al., 2000) dually infected with different strains of HIV-1 group M subtype B. Longitudinal analysis of another individual who was infected with two genetically distinct subtype B HIV-1 strains (approximately 15% divergent over the *env* C2-V5 coding region: Liu et al, 1997), identified a range of chimeric genome structures between the two original genotypes, which oscillated in frequency during the 56 month study period (Liu et al., 2002). Recombination between two distinct HIV-1 subtype B viral strains *in vivo* has also been demonstrated experimentally in a dually infected chimpanzee (Wei and Fultz, 1998).

1.6.4 Recombination between closely related HIV-1 strains

It is to be expected that recombination between individual HIV-1 variants that diverged during the course of a single infection would occur more frequently than between distinct members of the same subtype, or members of different subtypes. Closer sequence similarities are likely to facilitate the original generation of recombinants (Mikkelsen and Pedersen, 2000), and reduce potential incompatibilities between the newly juxtaposed sequences, leading to a corresponding increased likelihood that the recombinant genomes will produce viable virus. The probability

of any two variants present in an infected individual coinfecting a cell and becoming copackaged is also higher than the likelihood of this event occurring with two variants belonging to distinct populations. Yet the issues outlined above for detection of intrasubtype recombinants are compounded in the case of sequences that diversified within a single infected individual because of the very close sequence similarities and lack of pre-defined groups involved. Consequently detection of recombination between these variants has generally been inferred by very indirect means. For example a drastic reduction in *pol* sequence diversity identified in several patients undergoing antiretroviral therapy was not accompanied by a similar sequence bottleneck in *env*, suggesting a lack of genetic linkage between the two subgenomic regions (Dykes et al., 2000; Brown and Cleland, 1996). Similarly, soon after transmission the V3 region of gp120 appears to be subject to purifying selection (Pang et al., 1992; Zhu et al., 1993b; Wolinsky et al., 1992), and is less variable than other genomic regions such as V4/V5 (Wolinsky et al., 1992) and p17^{gag} (Zhang et al., 1993). One explanation for these findings is that strong early selection on V3 is accompanied by recombination among members of an initially diverse viral population (Zhang et al., 1993).

A study by Jung and colleagues has elegantly circumvented the problems associated with detection of recombination between closely related variants by obtaining proviral sequences from single infected cells (Jung et al., 2002). Using fluorescent *in situ* hybridisation, the number of proviruses per cell in single HIV-1 infected splenocytes from two patients was quantified. Remarkably, the proviral copy number ranged from one to eight per cell, with a mean of approximately 3.2. Amplification and sequencing of the V1/V2 region of *env* from single, laser-dissected interphase

nuclei revealed that those nuclei containing three or four proviruses harboured up to 34% amino acid differences in this genomic region. Moreover several of these proviral sequences appeared to be recombinants (Jung et al, 2002).

It is to be expected that recombination such as this (between HIV-1 variants that have diversified during a single infection) would provide a powerful mechanism for viral adaptation to the shifting selection pressures that characterise the natural course of HIV-1 infection.

1.7 Molecular determinants of tropism

Identification of the receptors that allow entry of HIV into host cells has provided many insights into the tropism and pathogenesis of this virus. The original discoveries and subsequent characterisation of HIV entry receptor usage have been extensively reviewed (Berger, 1997; Doms and Peipert, 1997; Rucker and Doms, 1998; Dimitrov et al., 1998; Clapham and McKnight, 2002; DSouza and Harden, 1996).

1.7.1 CD4

Binding of gp120 to CD4 is the first step required for HIV infection and this interaction is conserved among all primate lentiviruses. The initial evidence implicating CD4 as a receptor for HIV was the ability of anti-CD4 monoclonal antibodies to inhibit viral infection (Dalglish et al, 1984; Klatzmann et al., 1984b). Subsequently it was demonstrated that a complex of CD4 and the viral envelope could be immunoprecipitated from infected cells (McDougal et al., 1986), and that

expression of CD4 rendered previously resistant cells sensitive to HIV infection (Maddon et al, 1986).

CD4 is expressed predominantly on T-helper cells where its major role is to act as an accessory receptor enhancing inter-cellular connections during antigen presentation. CD4 interacts with components of the T-cell receptor (TCR) and binds the $\beta 2$ subunit of MHC class II molecules on antigen presenting cells, thus improving the avidity of associations between T-helper cells and MHC class II⁺ cells. It also serves more general roles as an adhesion molecule, a chemotactic receptor (responding to binding of IL-16: Cruikshank et al., 1998), and in the induction of cellular activation (Clapham and McKnight, 2002). In addition to T-helper cells dendritic cells, monocytes, macrophages, and microglia also express CD4 and can be infected by HIV (Gartner et al., 1986; Klatzmann et al, 1984a; Macatonia et al., 1990; Watkins et al., 1990). These cells do not express a TCR but their expression of CD4 is likely explained by additional functions of this molecule. A variety of other lymphoid cell types including CD8⁺ T lymphocytes, gamma-delta T lymphocytes, and natural killer cells have been shown to both express CD4 and contain HIV proviral DNA in some circumstances (Valentin et al., 2002; McBreen et al., 2001; Imlach et al., 2001; Imlach et al., 2003). However, the contribution to viral pathogenesis that infection of these alternative cell types represents is not clear. In summary the observed cellular distribution of HIV in the immune system closely correlates with expression patterns of the CD4 molecule.

1.7.2 Coreceptors

The discovery of CD4 as the primary receptor for HIV-1 explained several aspects of observed viral tropism. Yet it was also evident that an additional factor or receptor was also required for infection. Expression of CD4 alone was not sufficient to confer sensitivity to HIV-1 infection in some human cells (Chesebro et al., 1990) and most non-human cells (Maddon et al, 1986), despite apparently normal processing and expression of this molecule (Clapham et al., 1991). However, if non-human cells expressing human CD4 were fused with CD4-negative human cells then the resulting heterokaryon was permissive for HIV infection (Broder et al., 1993; Dragic et al., 1992), indicating that non-human cells lacked a requisite cofactor or coreceptor for infection to occur. Isolates of HIV-1 also exhibit two discrete cellular tropisms. Distinction between these viral phenotypes has variously been described as macrophage tropic (M-tropic) or T-cell tropic (T-tropic), depending upon the strains ability to infect macrophages or CD4⁺ T-cell lines in culture (Gartner et al, 1986); slow-low or rapid-high, as determined by replication rates in PBMCs (Asjo et al., 1986); and nonsyncytium-inducing (NSI) or syncytium-inducing (SI), defined by ability to mediate the formation of multinucleated giant cells in culture (Tersmette et al., 1988). With a small number of exceptions it became clear that these various definitions all described the same two biologically different viral types. One obvious explanation for this dichotomy was that the viral phenotypes represented strains that used different coreceptor molecules to enter cells (Alkhatib et al., 1996a), a hypothesis which was strengthened by the observation that determinants of differential tropism were mainly within the V3 loop of gp120 (Hwang et al., 1991).

The first identified HIV coreceptor, CXCR4 (previously called fusin, LESTR, or HUMSTR), was cloned in 1996 and facilitated entry of T-tropic/SI strains (Feng et al, 1996). Shortly afterwards CCR5, the coreceptor for M-tropic/NSI strains was identified simultaneously by several groups (Alkhatib et al, 1996b; Choe et al, 1996; Deng et al, 1996; Doranz et al, 1996; Dragic et al, 1996). These receptors are G protein-coupled seven-transmembrane domain receptors whose natural ligands are chemotactic cytokines (chemokines). The major role of chemokine receptors upon binding of their ligand is to mediate cellular responses involved in chemotaxis, cell migration, trafficking, and activation of leukocytes, although they also have a number of other functions, for example in growth regulation (reviewed in Baggiolini, 1998; Rollins, 1997). Correlation between the two previously identified viral phenotypes and use of either CXCR4 or CCR5 is far from absolute but does still hold true in many cases. To ease the previously tangled nomenclature HIV strains that use CXCR4 to enter CD4⁺ cells have been designated X4, those that use CCR5 designated R5, and dual tropic isolates that can use either coreceptor, R5X4 (Berger et al., 1998). CCR5 is expressed predominantly on activated memory T CD45RO⁺ cells, monocyte/macrophages, dendritic cells, and granulocyte precursors. Expression of CXCR4 is more widespread and includes naive CD45RA⁺ T cells, CD8⁺ T cells, monocyte/macrophages, and a number of other immune cell types (reviewed in Luster, 1998). A large number of additional coreceptor molecules have also been shown to facilitate infection of cell lines *in vitro* by specific strains of HIV and SIV (Table 1.1). These molecules are all chemokine receptors or closely related molecules with the same basic seven-transmembrane structure.

The importance of CCR5 as a coreceptor for HIV strains transmitted between individuals was demonstrated by the identification of a naturally occurring 32 bp deletion ($\Delta 32$) in the CCR5 gene, which leads to a premature stop codon and a truncated CCR5 protein that fails to reach the cell surface (Benkirane et al., 1997). Approximately 1% of Caucasians are homozygous for this allele (Dean et al., 1996) and these individuals are highly resistant although not immune to HIV infection via sexual (Dean et al, 1996), parenteral (Wilkinson et al., 1998), and vertical (Philpott et al., 1999) transmission routes. A number of other naturally occurring polymorphisms in chemokines or their receptors have also been identified, with varying consequences for susceptibility to HIV infection and disease progression (reviewed in O'Brien and Moore, 2000).

In addition to being preferentially transmitted R5 variants also predominate during the early asymptomatic phase and persist throughout infection. In approximately 50% of individuals infected with HIV-1 group M subtype B a “phenotypic switch” occurs late in the duration of infection, characterised by the appearance of SI variants able to use CXCR4 and often a more general broadening of viral coreceptor usage (Scarlati et al., 1997). This conversion is associated with rapid disease progression, rapid CD4⁺ T-cell decline, and reduced survival time after AIDS diagnosis (Koot et al., 1993).

Despite these associations a causative relationship between the change in viral tropism and disease progression has been difficult to establish, and the underlying factors leading to the phenotypic switch are presently unknown.

Coreceptor	Ligand	HIV-1	HIV-2	SIV	Reference for coreceptor use
CCR1	MIP-1 α , MIPF-1, MCP-3, RANTES	-*	+	+	(Bron et al., 1997; McKnight et al., 1998)
CCR2b	MCP-1, MCP-2, MCP-3	+	++	+	(Doranz et al., 1996)
CCR3	Eotaxin, eotaxin-2, MCP-3, MCP-4, RANTES	++	++	+	(Choe et al., 1996; Doranz et al., 1996)
CCR4	MDC, TARC, RANTES, MIP-1 α	-	+	-	(McKnight et al., 1998)
CCR5	MIP-1α, MIP-1β, RANTES, MCP-2	++++	++++	++++	(Alkhatib et al., 1996; Deng et al., 1996; Dragic et al., 1996)
CCR8	I-309	++	++	++	(Rucker et al., 1997)
CCR9	TECK	+	-	-	(Choe et al., 1998)
CXCR2	IL-8, NAP-2, ELR ⁺ , CXCs	-	+	NT	(Bron et al., 1997)
CXCR4	SDF-1	+++	+++	+	(Feng et al., 1996)
CX3CR1/V28	Fractalkine	+	+	+	(Reeves et al., 1997)
STRL-33/Bonzo/TYMSTR	CXCL16	+	+++	+++	(Alkhatib et al., 1997; Deng et al., 1997)
GPR1	?	+	++	+++	(Farzan et al., 1997)
GPR15/Bob	?	+	++	+++	(Farzan et al., 1997)
APJ	Apelin	+	+	+	(Choe et al., 1998; Edinger et al., 1998)
Chem R23	?	+	-	+	(Samson et al., 1998)
RDC1	?	+	+	+	(Shimizu et al., 2000)

Table 1.1. Coreceptors that support primate lentivirus infection of CD4⁺ cell lines *in vitro*. *Rarely or never used as a coreceptor (-); occasional use by a few isolates (+); used by 5-20% of isolates (++); frequent use by many isolates or by major subgroup (for example, CXCR4 use by R5X4 and X4 isolates, indicated in bold) (+++); major coreceptor used by predominant virus *in vivo* (for example, CCR5 use by HIV and SIV, indicated in bold) (++++). NT, not tested. Copied from (Clapham and McKnight, 2002).

Initial production of X4 variants is unlikely to be a limiting factor in the process as the small number of mutations required to acquire this phenotype coupled with the rapid rate of HIV evolution would be expected to result in the frequent production of new variants capable of exploiting receptors other than CCR5. It therefore appears that some *in vivo* selective constraint prevents the emergence of X4 variants until the latter stages of disease. The nature of this constraint is not clear although immune pressures such as neutralising antibodies, downregulation of CXCR4, or high levels of stromal derived factor-1 (SDF-1, the natural ligand for CXCR4) within the lymphoid tissue have all been suggested as potentially important factors (reviewed in Michael and Moore, 1999). Alterations in the predominant cytokine or chemokine profiles may therefore be an important feature in the change in receptor usage (reviewed in Kinter et al., 2000). For example interleukin-4 has been shown to decrease expression of CCR5 while simultaneously upregulating CXCR4 expression in CD4⁺ T-cells, thereby providing a selective advantage for X4 over R5 variants (Valentin et al., 1998).

All HIV-1 isolates identified to date can use either CCR5 or CXCR4 to enter cells, and strains that utilise additional receptors usually do so with much less efficiency than via these major coreceptors. Methods for characterising the entry requirements of a particular viral isolate often involve expression of plasmids encoding CD4 and a coreceptor “out of context” in cell lines that do not naturally express these molecules. Several factors must be taken into account when interpreting tropism data obtained in this manner and applying the results to *in vivo* situations. Firstly, this approach may lead to the expression of receptor combinations that are not present *in vivo*, for example APJ is not naturally expressed at detectable levels on CD4⁺ cells (Puffer et

al., 2000; Choe et al., 2000) therefore its significance as an HIV coreceptor is questionable. Secondly, cell line expression systems appear to produce cell surface receptor molecules at a higher density and in a conformation more amenable for HIV-1 entry than primary cell types (Doms, 2000; Clapham and McKnight, 2002). The biological importance of coreceptors other than CCR5 and CXCR4 is therefore unclear. There is some evidence for infection of thymocytes via CCR8 (Lee et al., 2000), and that STRL-33 is expressed at sufficiently high levels to support viral infection of a subset of CD4⁺ T-cells (Sharron et al., 2000). CCR3 has also been shown to facilitate infection of microglial cells in some primary human foetal brain cultures (He et al., 1997), although this does not appear to be a general mechanism for microglial infection (Albright et al., 1999; Ghorpade et al., 1998b). Despite these possible exceptions there is an overall lack of evidence for frequent *in vivo* use of coreceptor molecules other than CCR5 and CXCR4 by HIV-1.

Considerable interest does still exist in identifying receptors that are expressed at sufficient levels *in vivo* to potentially allow their use as coreceptors for HIV. This attention is driven in part by the fact that chemokine receptors are attractive targets for therapeutic interventions to prevent entry of HIV into cells, with the apparent health of CCR5 Δ 32 homozygous individuals suggesting that this receptor may be aggressively targeted with minimal side effects. Exerting such strong selective pressure may result in altered patterns of HIV coreceptor usage with unpredictable consequences for viral cellular tropism. Therapies targeted at specifically inhibiting the use of CCR5 as a coreceptor may drive the virus to use CXCR4 and possibly instigate the clinical decline associated with this phenotypic switch (reviewed in Michael and Moore, 1999), or promote the evolution of variants that can utilise other

coreceptor molecules which are rarely or never used at present. An example of the unusual coreceptor use that may result from selective pressure at the level of viral entry is provided by SIV strains isolated from red-capped mangabeys in Gabon. In contrast to most SIV strains, these isolates use CCR2b as their major coreceptor. The likely explanation for this altered phenotype is that the majority of red-capped mangabeys are homozygous for a defective CCR5 gene (Chen et al., 1998).

1.7.3 Further factors involved in viral entry

A number of biological differences between HIV-1 isolates are manifest at the level of viral entry yet cannot be simply explained by the presence or absence of particular receptors on target cells. For example it is not clear why R5 variants are almost exclusively transmitted and R5X4 variants infrequently infect new individuals by the same route. Macrophages express sufficient levels of CD4, CCR5, and CXCR4 to allow HIV-1 entry (Lee et al., 1999b) but only a subset of variants able to use CXCR4 can infect these cells (generally primary virus isolates as opposed to T-cell line adapted virus) (Simmons et al., 1998; Yi et al., 1998). There are also several other examples where a virus strain has proven unable to enter cell lines or primary cells that appear to express the required receptors (Bazan et al., 1998; Dittmar et al., 1997), and where an increase in ability to direct membrane fusion is associated with small changes in *env* without concomitant changes in the receptors used (Doms, 2000). Biological differences such as these are likely to be determined by more subtle interactions between Env and the cellular receptors used for HIV entry than simple specificity for a certain coreceptor molecule.

Numerous factors may affect the ability of a particular viral strain to enter cells even if the appropriate receptors are present. The density of CD4 and coreceptors varies on different cell types (Lee et al, 1999b), and it is clear that HIV-1 isolates can differ greatly in their affinity for both of these molecules (Kozak et al., 1997; Kabat et al., 1994). CD4 and coreceptor density requirements also appear to be interrelated. For example one study demonstrated that when CD4 is expressed at high levels then relatively low levels of CCR5 are necessary to allow maximum viral infection and if the expression of CD4 is reduced then a high level of CCR5 is required (Platt et al., 1998). Thus cells that express levels of receptor molecules below the threshold required by a particular viral strain will be refractory to infection. Coreceptor molecules also exist in multiple antigenic conformations, and this can vary in a cell type specific fashion (Baribaud et al., 2001; Lee et al., 1999a). Such differences may be due to interactions with CD4 on the cell surface as has been described for CCR5 (Wu et al., 1996) or dimerisation of chemokine receptors (Lapham et al., 1999). The full consequences of these multiple conformational states in ability to facilitate infection by different viral strains have yet to be determined.

The specific regions of coreceptor molecules used by different viral strains also vary (reviewed in Clapham and McKnight, 2002). Chemokine receptors form pores in the cellular membrane with four domains exposed on the cell surface, the N terminus and three extracellular loops (ECL1, ECL2, and ECL3) (Palczewski et al., 2000). The N terminus of CCR5 is important for the entry of R5 variants (Hill et al, 1998), although generation of chimeric human/mouse CCR5 receptors has demonstrated that specific HIV-1 isolates differ in their interactions with CCR5 (Picard et al., 1997). In the case of X4 variants the ECL2 domain of CXCR4 is critical for

coreceptor function (Lu et al, 1997). X4 strains also vary in their usage of CXCR4, with different isolates dependent upon distinct ECL2 residues for coreceptor activity (Brelot et al., 1999). This capacity for different Envs to utilise distinct residues on coreceptor molecules presents adaptive possibilities, for example in allowing the evolution of variants with a reduced sensitivity to inhibition by neutralising antibodies or chemokine receptor ligands.

Interaction of gp120 with both CCR5 and CXCR4 can lead to the activation of intracellular signalling pathways (Davis et al., 1997; Weissman et al., 1997). It is possible that this signalling helps to prepare the cellular environment for optimal viral replication and generation of progeny virions. Signalling is not a prerequisite for productive viral infection to occur (Aramori et al., 1997; Gosling et al., 1997), although it has been reported that the ability of viruses to pass through early post-fusion events in macrophages correlates with their capacity to signal via CCR5 (Arthos et al., 2000). gp120 shed from infected cells could also induce chemotaxis in nearby uninfected lymphocytes leading to their recruitment and subsequent infection. A further consequence of cell signalling by soluble gp120 is that its attachment to receptors on uninfected bystander cells can lead to toxic effects, even if the cells express only CCR5 or CXCR4 and are thus largely refractory to direct infection by HIV. This effect is thought to be important for HIV pathogenesis in the brain where chemokine receptors are expressed on numerous cell types whose normal function is strongly affected by the presence of HIV but are only infrequently or never infected (reviewed in Gabuzda and Wang, 2000).

1.7.4 Envelope determinants of HIV-1 cellular tropism and coreceptor use

Envelope sequences obtained from primary HIV-1 isolates within single infected individuals are often extremely heterogeneous and can be up to 10% divergent at the nucleotide level. This high level of genetic variability demonstrates a remarkable plasticity and tolerance to mutations within the *env* gene and has important implications for viral pathogenesis. Viral replication in specific cell types can be restricted at a number of entry and postentry steps, but early studies indicated that the restriction to HIV replication in non-permissive CD4⁺ cells was primarily at the level of viral entry (Cheng Mayer et al., 1990; Kim et al., 1990; O'Brien et al., 1990). The envelope protein (Env) mediates binding to a target cell followed by fusion of the viral and host cell membranes. It was therefore unsurprising that the principal determinants of both coreceptor usage and macrophage or T-cell tropism in culture were found to be within Env. Further alterations in *env* may lead to more subtle *in vivo* adaptations allowing the expansion of virus into new cell or tissue types, for example by affecting receptor affinity, immune evasion, or resistance to chemokine inhibition of infection.

Generation of chimeric envelopes from M- and T-tropic isolates led to the identification of the V3 loop as one of the primary determinants of viral tropism (Chesebro et al., 1991; Hwang et al, 1991; Shioda et al., 1991). In the context of certain *env* sequences even a single amino acid change in V3 is sufficient to alter cellular tropism (Takeuchi et al., 1991). The most comprehensively characterised of these mutations is the presence of positively charged amino acids at positions 11 or 25 situated on either side of the crest region of the V3 loop which closely correlates

to a SI viral phenotype (Fouchier et al., 1992). However, many mutations in the V3 loop that alter tropism have strain-specific effects and no single residue or amino acid motif is crucial for the phenomenon. In general a low overall positive charge in the V3 loop and few mutations as compared with the subtype B consensus sequence is indicative of M-tropic virus (Donaldson et al., 1994a). This is consistent with longitudinal studies of HIV infected individuals where an increase in the net positive charge in the V3 loop is associated to conversion to T-tropism (de Jong et al., 1992a; de Jong et al., 1992b; Shioda et al., 1992; Fouchier et al., 1992). The V3 loop is also important in determining coreceptor usage, with the same positively charged amino acids that confer a SI phenotype also correlating with use of CXCR4 (Speck et al., 1997).

The V1/V2 region has similarly been shown to effect cellular tropism in many virus strains. During receptor binding the V1/V2 loops appear to act co-operatively with V3 and sequence alterations in V1/V2 can modulate both cellular tropism and coreceptor usage in a V3 specific manner (Carrillo and Ratner, 1996; Shioda et al., 1991; Koito et al., 1994; Cho et al., 1998; Ross and Cullen, 1998). A link between positively charged residues in V1/V2 and T-tropism has been noted (Groenink et al., 1993) and in some cases a transient increase in the length of V2 is associated with an *in vivo* switch from NSI to an SI viral phenotype (Fouchier et al., 1995a). Several mutations that can affect cellular tropism in this region are also N-linked glycosylation sites (Koito et al., 1995; Carrillo and Ratner, 1996). However, correlation of these mutations with viral tropism is far from absolute, and V1/V2 does not appear to have as critical a role in envelope function as V3. For example

while deletion of V3 results in the production of non-infectious virus (Wyatt et al, 1993) deletion of V1/V2 does not abrogate viral infectivity (Cao et al., 1997).

A crystal structure of gp120 in a complex with CD4 and the neutralising CD4 induced antibody 17b has been described (Kwong et al, 1998). This gp120 had most of the V1/V2 and V3 loops deleted and lacked 90% of the carbohydrate present in the natural protein, but these alterations did not appear to affect the global architecture of the molecule. Information gained from this structure allowed the sites on gp120 that interact with coreceptors after CD4 binding to be inferred. They include the V1/V2 and V3 loops as well as a conserved region composed of β -strands situated between these loops and residues in C4, known as the bridging sheet. Thus the regions of Env previously determined to influence tropism do so (at least in part) by direct interaction with coreceptor molecules.

Further insights into the specific mechanisms by which Env sequences can affect chemokine receptor usage have been gained by characterising the regions of coreceptor molecules that are important for binding different gp120 molecules (Clapham and McKnight, 2002). The N terminus of CCR5 (and several other coreceptors) is negatively charged due to four (potentially) sulphated tyrosine residues and three acidic amino acids (Farzan et al., 1998). These residues are important for coreceptor function and may aid electrostatic interactions with basic amino acids in the bridging sheet of R5 gp120 molecules (Kwong et al, 1998). Electrostatic interactions may also help explain the observed association of CXCR4 usage with a more highly positively charged V3 loop. The E2 loop of CXCR4 is critical for X4 virus entry (Lu et al, 1997; Brelot et al, 1999) and contains five

negatively charged residues that may interact more readily with a positively charged V3 loop. However, replacement of all five of these acidic amino acids with alanines does not eliminate coreceptor function (Wang et al., 1998), therefore the specificity of the gp120-coreceptor interaction may be aided but is not completely governed by electrostatic interactions of this type.

1.7.5 CD4 and coreceptor independent interactions

It is now well established that most HIV-1 primary virus isolates require CD4 and a member of the seven-transmembrane chemokine receptor family as entry receptors for fusion with host cells. However a number of additional cell surface molecules also bind to gp120 with high affinity (reviewed in Clapham and McKnight, 2002). These molecules may therefore function as attachment receptors, facilitating viral adsorption to the cell surface without a direct role in membrane fusion. Interactions auxiliary to direct gp120-CD4 binding may be necessary for infection of cell types such as dendritic cells, macrophages, and microglia, where *in vivo* CD4 expression is much lower than found in CD4⁺ T-cells (Sonza et al., 1995; Dick et al., 1997). Infection of CD4-negative cell types *in vivo* has also been poorly characterised at the molecular level, and adsorption of virus to these cells may well require a stronger affinity than is achieved between gp120 and coreceptor molecules alone.

The glycolipid galactocerebroside (galactosylceramide, Gal-C) and its sulphated derivative (sulphatide) both bind gp120 with high affinity (Fantini et al., 1993; Harouse et al., 1991). These receptors are expressed on macrophages (Seddiki et al., 1996; Seddiki et al., 1994), neural and glial cell lines (Harouse et al, 1991), and colonic epithelial cell lines (Fantini et al, 1993). When a coreceptor is present then

Gal-C can facilitate low-level infection by particular HIV-1 strains in a CD4-independent manner (Delezay et al., 1997) leading to the suggestion that these molecules are important factors in HIV infection of CD4-negative cells in the brain and colon (Harouse et al., 1995). The molecular mechanism of this infection is unclear, as CD4 binding is usually required to produce the conformational changes in Env necessary for coreceptor binding and membrane fusion. Yet a number of CD4-negative cell types do appear to support some level of HIV-1 infection *in vivo* and *in vitro*, thus alternative interactions allowing CD4-independent adsorption of virus to host cells may be an important first step in the infection process.

gp120 is heavily glycosylated and CD4-independent attachment of virions may therefore also be mediated through carbohydrate groups present on this molecule associating with lectin-like domains on cellular receptors (Larkin et al., 1989). The glycosaminoglycan heparan sulphate was found to be important in HIV-1 binding to both CD4⁺ T-cell lines (Roderiquez et al., 1995; Patel et al., 1993), and HeLa cells with or without CD4 expression (Mondor et al., 1998). This interaction appears to be mainly between positively charged residues in the V3 loop of gp120 with negative sulphate groups on glycosaminoglycans, thus the interaction is stronger with X4 and R5X4 viruses than R5 viruses because the latter generally have a lower overall positive charge in the V3 loop (Moulard et al., 2000). Dendritic cell-specific intercellular adhesion molecule-grabbing nonintegrin (DC-SIGN) is a C-type lectin cell surface protein with a potentially important biological role in HIV binding (reviewed in Pohlmann et al., 2001a). This molecule is expressed on some dendritic cell populations (Geijtenbeek et al., 2000a; Geijtenbeek et al., 2000b) and had previously been shown to bind gp120 with high affinity (Curtis et al., 1992). The presence of

DC-SIGN on dendritic cells provides a model not only for HIV-1 infection of this cell type despite low levels of CD4 expression, but also for transport of HIV from mucosal membranes to lymph nodes upon original exposure to virus. Following mucosal transmission the first cells likely to be encountered by HIV are immature dendritic cells such as Langerhans cells. During this first contact the virus may either directly infect dendritic cells or become trapped on the cell surface by DC-SIGN and carried to a lymph node where it could be transferred to T-cells during antigen presentation, thus allowing a rapid amplification of virus and enhanced subsequent systemic dissemination. A homologue of DC-SIGN (DC-SIGNR) binds HIV in a similar manner but is expressed on endothelial cells (Pohlmann et al., 2001b).

Further interactions between HIV virions and the host cell may be mediated not directly through gp120 but rather via association of target cell surface components with adhesion molecules incorporated into the viral envelope during budding from the cytoplasmic membrane (Ugolini et al., 1999). The observation that CD4 alone, but not CD4 associated with its natural intracellular ligand (tyrosine kinase p56lck) is incorporated into virions (Henriksson and Bosch, 1998) has led to the suggestion that such incorporations may be not completely random, and that there may be a some level of selectivity in the process (reviewed in Tremblay et al., 1998). Leukocyte function-associated antigen 1 (LFA-1) specific monoclonal antibodies (mAbs) were found to interfere with HIV induced syncytium formation *in vitro* (Hildreth and Orentas, 1989), and anti-intercellular adhesion molecule 3 (ICAM-3) mAbs reduced infectivity of cell-free virus (Sommerfelt and Asjo, 1995). Producing HIV in cells transfected with genes coding for specific adhesion molecules provides an indication of their potential importance in viral attachment. For example the presence of HLA-

DR in producer cells increased HIV infectivity for CD4⁺ cells by approximately twofold (Cantin et al., 1997). While up to a tenfold increase in infectivity was observed when intercellular adhesion molecule 1 (ICAM-1) was present in producer cells and progeny virus was used to infect LFA-1⁺ cells (Paquette et al., 1998; Fortin et al., 1999). Thus recognition of adhesion molecules present in the viral envelope by their cognate receptors anchored on the target cell surface may enhance viral infectivity in both a producer and target cell type specific manner.

Therefore the initial adsorption of HIV to host cells is governed by many interrelated factors that could affect the *in vivo* cellular tropism of HIV-1 viral variants.

1.8 Compartmentalisation of HIV-1

Soon after transmission HIV-1 disseminates widely throughout the body. At this point viral sequences in peripheral blood and tissues are closely related (Zhang et al., 2002), and the virus population as a whole is much more genetically homogeneous than is found at later timepoints (Zhu et al., 1993a; Shankarappa et al., 1998; McNearney et al., 1992). As infection progresses distinct viral subpopulations are frequently identified in different anatomical sites, most evidently when the infected individual has progressed to AIDS before death. Compartmentalisation may be detected by sequence differences, differential quantities of viral nucleic acids, or the presence of phenotypically distinct viral variants (for example differing in cellular tropism or susceptibility to antiviral agents) in different sites. There are several possible explanations for the emergence of divergent viral sub-populations. These include physical barriers between compartments, selective trafficking of infected cells, localised selection pressures (such as the magnitude and specific manifestation

of immune responses), and both the type and availability of susceptible cells. Anatomical sites found in some cases to harbour viral variants that are phylogenetically distinct from those in peripheral blood include the brain (Donaldson et al, 1994a; Wong et al., 1997; Korber et al., 1994a), cerebrospinal fluid (Steuler et al., 1992; Tang et al., 2000), genitourinary tract (Delwart et al., 1998; Zhu et al., 1996; Ping et al., 2000; Poss et al., 1998), colon (Wang et al., 2001), bone marrow (Voulgaropoulou et al., 1999), lung (Wang et al, 2001; Nakata et al., 1995; Vantwout et al., 1998; Itescu et al., 1994), and kidney (Marras et al., 2002). Proviral sequences in the peripheral blood, spleen, and lymph node are usually closely related (Ball et al., 1994; Donaldson et al, 1994a), and are therefore considered to represent a single "lymphoid" compartment. However sequence comparisons between plasma viral RNA and the proviral DNA present in PBMCs often reveal genetic differences, with variants detected in viral RNA generally not observed in proviral DNA until later timepoints (Simmonds et al., 1991; Ball et al, 1994; Wei et al, 1995).

Elucidating the viral and host factors that facilitate or drive the evolution of these distinct viral sub-populations is likely to enhance our understanding of various aspects of HIV biology, including sexual transmission and neuropathogenesis. It will also help define and characterise sanctuary sites in patients undergoing antiretroviral therapy, thereby providing specific bioavailability objectives for the development of antiviral agents.

1.9 HIV-1 and the brain

Several cell types in the central nervous system (CNS) are susceptible to infection by HIV-1, and in late stage AIDS the CNS often constitutes the largest reservoir of

replicating virus outwith the immune system. The presence of HIV within the CNS frequently leads to neurological symptoms and a characteristic pathological appearance.

1.9.1 Neurological symptoms and pathology associated with HIV-1 infection

Approximately 30-60% of HIV-1 infected individuals experience neurological or psychiatric symptoms, which may show gradual onset and become more severe with disease progression. These can include mental slowness, confusion, headaches, memory loss, behavioural changes, and motor signs (such as unsteady gait and tremor), that eventually culminate in dementia in approximately 10% of patients with AIDS (Bell, 1998; Gray et al., 1996; Epstein and Gendelman, 1993). These symptoms together with a progressive cognitive decline have variously been termed AIDS related dementia (ARD), the AIDS dementia complex (ADC), or HIV associated dementia (HAD). HAD represents a significant independent risk factor for death due to AIDS (Lipton and Gendelman, 1995), and has been classified as an AIDS defining illness. In developed countries the use of HAART has resulted in a decrease in the prevalence of HAD in patients with advanced HIV disease and low CD4 counts from approximately 20-30% in the early 1990s to the current prevalence of around 10% (Mcarthur et al., 1993; Ferrando et al., 1998; Sacktor et al., 2001). However, the proportion of new cases of HAD in individuals with a CD4 count greater than 200/ μ l may be increasing (Sacktor et al, 2001). There is also evidence that HAART does not completely protect against or cause reversal of HAD (Dore et al., 1999; Major et al., 2000). Moreover the CNS may constitute a sanctuary site for HIV and allow continued viral replication despite effective suppression by HAART

elsewhere in the body. Consequently as the lifespan of HIV-1 infected individuals is prolonged by the use of antiviral therapies and treatment of opportunistic infections it is possible that the incidence of HAD may actually increase. HAD therefore remains a significant cause of morbidity and mortality in HIV-1 infected individuals.

Neuropathological studies have demonstrated abnormalities in the CNS of 80-100% of AIDS patients (reviewed in Gray et al, 1996; Bell, 1998). In pre-AIDS cases pathological changes are fairly limited but may sometimes include an inflammatory T cell reaction, an increased number of microglial cells, upregulation of major histocompatibility complex class II antigens, localised production of cytokines, and other changes also consistent with an immunological process (reviewed in Gray et al, 1996). Similar assessments of AIDS autopsies have revealed more profound neuropathological changes, which can include the presence of opportunistic infections (such as cytomegalovirus, *Toxoplasma gondii*, and JC virus) and lymphomas. HIV encephalitis (HIVE) is a direct effect of HIV-1 within the CNS and is observed in 10-50% of AIDS autopsy series. HIVE is characterised by the presence of multinucleated giant cells and/or the detection of HIV-1 proteins or nucleic acids in brain parenchymal cells. A definitive diagnosis of HIVE requires the direct identification of viral antigens or nucleic acids, and can be made in the absence of multinucleated giant cells. HIVE is often associated with inflammatory (immune cell mediated) and degenerative (myelin loss and axonal damage) white matter damage. Other neuropathology in AIDS cases can include neuronal loss, astrocytosis, inflammatory infiltrates, and apoptosis in a number of different cell types (reviewed in Bell, 1998; Epstein and Gendelman, 1993). However, there is no exact correlation

between the clinical expression of HAD and any neuropathological changes observed at autopsy (reviewed in Bell, 1998).

1.9.2 Timing of HIV-1 infection of the brain

The blood-brain-barrier (BBB) is composed of brain microvascular endothelial cells connected to astrocyte foot processes by tight junctions. Under normal circumstances the BBB restricts the movement of cells, proteins, and ions between the circulation and the CNS. HIV-1 probably enters the CNS via infected monocytes or possibly lymphocytes trafficking across the BBB from the peripheral blood (Price, 1996). This event can occur soon after transmission, perhaps before seroconversion (Davis et al., 1992; Palmer et al., 1994). However, while infectious HIV-1 can usually be recovered from the cerebrospinal fluid at all timepoints of infection (Chiodi et al., 1988; Chiodi et al., 1992), only very low or undetectable levels of HIV-1 antigens and proviral DNA are usually observed in the brains of asymptomatic individuals (Donaldson et al., 1994a; Donaldson et al., 1994b; Gray et al., 1992). One possible interpretation of these observations is that early invasion of the brain by HIV-1 seeds a latent, low-level infection, which subsequently reactivates with progression to AIDS. Alternatively, maintenance of HIV-1 in the brain at such low levels may not be plausible, considering that many years often separate acute infection from the development of HIV-1 associated neurocognitive disorder (HIV-1 associated neurocognitive disorder) (Gartner and Liu, 2002; Gartner, 2000). This scenario would suggest that infection of the brain would need to be re-established at a later timepoint, possibly with the onset of AIDS, and that infected cells trafficking from elsewhere in the body are the proximate source of virus in HIV-1 associated neurocognitive disorder (as opposed to reactivation of virus already present at this site). Distinguishing between these two models is important for furthering our basic understanding of HIV-1 infection in the

brain, and in assessing the significance of the CNS as a sanctuary site during HAART. The relative merits and implications of these models will therefore be discussed further (Chapter 3)

1.9.3 Infected cell types in the brain

The predominant cell types infected by HIV-1 in the CNS are monocytes, macrophages, and microglial cells (Koenig et al., 1986; Wiley et al., 1986). It is thought that HIV-1 Env-directed fusion of these infected cells with uninfected neighbouring cells gives rise to the characteristic syncytia that often accompany HIV-1. Some reports have suggested that HIV-1 infects a wider range of cell types in the CNS including astrocytes (Cheng Mayer et al., 1987; Chiodi et al., 1987), oligodendrocytes (Bagasra et al., 1996; Gyorkey et al., 1987), neurones (Pumarola Sune et al., 1987; Bagasra et al., 1996), and microvascular endothelial cells (MVECs) (Wiley et al., 1986; Pumarola Sune et al., 1987). The current consensus opinion is that oligodendrocytes and neurones are not infected by HIV-1. Infection of astrocytes has been demonstrated both *in vivo* and *in vitro* by a number of techniques, but viral replication appears to be restricted, with often only accessory gene products expressed (reviewed in Brack-Werner, 1999). The nature of this restriction is incompletely defined but has variously been ascribed to a block in Rev function (Ludwig et al., 1999; Neumann et al., 1995), defects in envelope processing (Shahabuddin et al., 1996), and inefficient translation of structural gene products (Gorry et al., 1999). In contrast, a study of primary human foetal astrocytic infection by HIV-1 pseudotyped with Envs of amphotropic murine leukemia virus or vesicular stomatitis virus showed a highly productive infection of these cells, suggesting astrocyte infection by HIV-1 is restricted at the level of viral entry (Canki et al.,

2001). Astrocytes do not express CD4, but subpopulations of astrocytes can express CCR5 and CXCR4 (reviewed in Hesselgesser and Horuk, 1999; Gabuzda and Wang, 2000) . Therefore infection of astrocytes appears to be via a CD4-independent mechanism. While CD4-independent X4 and R5 HIV-1 strains have been isolated after viral passage in tissue culture (Kolchinsky et al., 1999; Dumonceaux et al., 1998), and a number of CD4-negative cell types are sometimes found to be susceptible to HIV-1 infection both *in vivo* and *in vitro* (Speck et al., 1999) and references therein), in the vast majority of cases infection of human cells with naturally occurring HIV-1 isolates requires the presence of CD4. Consequently the specific mechanism by which astrocytes are infected remains poorly defined. MVECs also do not express CD4, but do express CCR5, CXCR4, DC-SIGN, and L-SIGN (Mukhtar et al., 2002). Our current understanding of HIV-1 infection of MVECs is therefore incomplete for the same reasons described for astrocyte infection. In the SIV/macaque animal model, which shares many common features with HIV-1 infection in humans, neurovirulent strains of SIV have been identified. One such strain (SIV/17E-Fr) was found to infect MVECs in a CD4-independent manner (Edinger et al., 1997). The occurrence of similar CD4-independent infection of MVECs by HIV-1 is somewhat controversial. However, if true then this would provide an easy portal for viral entry into the CNS, and may result in disruption of BBB function, thereby contributing to the development of HAD.

1.9.4 HIV-1 induced damage in the CNS

HIV-1 does not appear to infect neurones. Neurodegeneration in HIV-1 infection is therefore thought to be largely the result of indirect mechanisms, probably instigated by factors released from activated or infected monocyte derived macrophages and

microglial cells within the brain. One possible mechanism for neuronal damage is the shedding of viral proteins by infected cells. Many HIV-1 gene products have been shown to have toxic effects upon neurones and other CNS cell types (Kanmogne et al., 2002; Shi et al., 1998; Adamson et al., 1999; Sabatier et al., 1991). Perhaps the most likely candidate viral protein is gp120, which is frequently shed by infected cells. Both soluble and virion-associated gp120 can trigger signal transduction pathways by binding receptors such as CXCR4 and CCR5 present on neurones and other cell types in the CNS, thereby affecting cellular functions and inducing apoptotic death (Kaul et al., 2001). A further pathway to neuronal damage and apoptosis is secretion of inflammatory cytokines and toxic factors by activated or infected cells. Factors released in this manner are thought to establish a complex cascade of inter- and intra-cellular signal transduction pathways and networks, ultimately resulting in neuronal death (reviewed in Kaul et al, 2001). Molecules implicated in this process include glutamate, quinolinic acid, platelet aggregating factor, α - and β -chemokines, interleukin-1 β , tumour necrosis factor- α , and free radicals (Lipton and Gendelman, 1995).

1.9.5 HIV-1 nucleotide sequences in the brain

Little is known about the sequence relationships between viral variants present in brain and lymphoid tissue during the asymptomatic period of infection. This is primarily because the low HIV-1 proviral loads detected in the brain at this point could be accounted for by the presence of infected cells in residual blood within the tissue samples (Donaldson et al, 1994a; Donaldson et al, 1994b; Gray et al, 1992). The availability of tissue specimens representing this stage of infection is also relatively limited. In contrast, a large number of investigations have focussed upon

determining the sequence relationships of virus from the brain and other compartments in individuals who have died of complications associated with AIDS. These comprise a diverse body of work with many differences in sample materials and study designs, but some broad patterns have emerged. The majority of investigations have focussed upon areas of the *env* gene (usually including the V3 region) as this is generally considered to be the primary determinant of HIV-1 cellular tropism. Many studies have found that virus in the brain of individuals with HIV is phylogenetically distinct from that in the lymphoid compartment (Epstein et al., 1991b; Epstein et al., 1991a; Haggerty and Stevenson, 1991; Korber et al, 1994a; Pang et al., 1991; Chang et al., 1998; Donaldson et al, 1994a; Gartner et al., 1997; Gorry et al., 2001; Hughes et al., 1997a; Shapshak et al., 1999), although this is not necessarily the case for every subgenomic region or study subject analysed (Hughes et al., 1997b; Korber et al, 1994a). Inferences of coreceptor preference and cellular tropism based on V3 sequences in brain almost invariably predict a CCR5-using, macrophage tropic viral phenotype (Korber et al, 1994a; Wang et al, 2001; Donaldson et al, 1994a; Chen et al., 2000). Viral variants similar to those present in lymphoid tissue are sometimes sampled at lower levels in the brain, perhaps representing a small amount of peripheral blood or recently infiltrating lymphocytes/macrophages in tissue samples (Korber et al, 1994a; Gatanaga et al., 1999; Hughes et al, 1997b). The overall viral sequence diversity within brain can also be greater than is found within lymphoid tissue (Hughes et al, 1997b). Furthermore, sequence similarities have also been noted between variants in brain and those present in other macrophage-rich compartments such as the lung and colon (Wang et al, 2001). Compartmentalisation between brain and lymphoid variants is

sometimes reflected in other regions of the viral genome such as LTR (Ross et al., 2001), and *pol* (Wong et al, 1997), although only a limited number of studies have focussed on subgenomic regions outwith *env*. Many factors can lead to the development of distinct viral subpopulations including adaptation to a specific environment, physical isolation, and differences in replication rates. The relative contributions of each of these factors are largely unknown for HIV-1 in the CNS.

It has been suggested that specific viral *env* sequences are associated with the occurrence of dementia (Power et al., 1994; Power et al., 1998). Other studies have failed to find such correlations between the presence of particular amino acid signature sequences and the clinical presentation of HAD (Reddy et al., 1996; Di Stefano et al., 1996). In general the only consistent finding between studies of HIV-1 in the brain is the presence of a low overall positive charge in the V3 loop, broadly (although not always) consistent with macrophage tropism. Hence there does not appear to be any clear set of mutations that lead to a “neurotropic” or “neurovirulent” viral phenotype, although the question of whether such a phenotype may be converged upon by many different constellations of mutations remains open.

1.9.6 Phenotype of HIV-1 in the brain

The presence of distinct genetic variants within the CNS gives rise to the possibility that adaptive changes may occur within this virus population. For example local immune responses and the phenotype of available target cells in the CNS microenvironment may select for the outgrowth of particular viral variants. Consistent with inferences based upon partial *env* nucleotide sequences, most HIV-1 variants obtained from brain tissue have a macrophage tropic phenotype (Cheng-

Mayer and Levy, 1988; Li et al., 1991), and require both CD4 and CCR5 to enter microglial cells (the resident CNS macrophages) (Albright et al, 1999; He et al, 1997; Shieh et al., 1998). HIV-1 infection of the CNS alone is not sufficient to cause dementia (Bell, 1998; Power et al, 1994). Likewise, the formation of syncytia and induction of apoptosis within the CNS does not occur in all individuals with evidence of HIV-1 at this site (Bell, 1998). Moreover during *in vitro* cell culture infection experiments only a subset of viruses produce syncytia in microglial cultures (Strizki et al., 1996; Watkins et al, 1990), or induce neuronal apoptosis in primary brain cultures (Ohagen et al., 1999; Power et al, 1998). These observations have led to the proposal that “neurotropic” HIV-1 strains (as defined by ability to replicate in microglial cells) are not necessarily “neurovirulent” i.e. do not cause the release of factors that induce neuronal apoptosis (Power et al, 1998; Gorry et al., 2002a). No consistent amino acid changes have been identified that can be used to predict whether a particular R5 isolate will be able to replicate efficiently and/or induce syncytia in microglial culture, or cause apoptosis in neighbouring neurones. Therefore the determinants in Env that lead to these phenotypes appear to be conformationally complex and context dependent.

Microglial cells and macrophages express lower levels of CD4 than is present on CD4⁺ T cells in the peripheral blood (Dick et al, 1997; Lewin et al., 1996). A reduced dependence on CD4 for viral entry would therefore be a plausible adaptation for enhanced HIV-1 infection of cells in the brain. With respect to this issue an informative series of experiments have been performed in the laboratory of González-Scarano and colleagues (Strizki et al, 1996; Shieh et al., 2000; Martin et al., 2001). During these experiments 27 primary virus isolates were obtained from

acutely infected individuals and assessed for their ability to replicate in either monocyte derived macrophages (MDM) or microglial cells (Strizki et al, 1996). While most isolates replicated equally well in both, several strains replicated preferentially in one of the two cell types. Serial passage of one isolate (HIV-1_{BORI-15}) in microglial cultures led to a 1000-fold increase in peak virus production (Strizki et al, 1996). HIV-1_{BORI-15} replicated to a higher titre in microglia and MDM than in PBMCs and induced prominent syncytia, predominantly in microglia. This altered phenotype was not due to a change in receptors used for virus entry, as the parental and microglia-adapted virus each required the presence of both CD4 and CCR5 (Shieh et al, 2000). Generation of chimeric viruses revealed that the determinants for the new syncytium-inducing phenotype were within Env, and sequence analysis of the parental and microglia-adapted envelopes showed differences in eight amino acid positions. Four amino acid differences were within the V1/V2 region of *env*, and two of these eliminated potential N-linked glycosylation sites (Shieh et al, 2000). Further characterisation of this strain demonstrated that the alterations in V1/V2 permitted infection of cells expressing lower levels of CD4 (Strizki et al, 1996; Shieh et al, 2000). Infection experiments using cells expressing chimeric CCR5/CCR2b coreceptor molecules established that differences between the microglia-adapted and parental isolates also led to alterations in specific interactions with the coreceptor molecule during viral entry. *env*-pseudotyped virus containing either the V1/V2 region of HIV-1_{BORI-15} in the context of the parental isolate (rV1/V2), or the entire *env* of HIV-1_{BORI-15} (rB15) led to more efficient infection of cells expressing CCR5 with the third extracellular loop (ECL3) substituted for that of CCR2b than did the parental *env* (rBORI). Differences in the dependence upon ECL3 were also seen

between rV1/V2 and rB15, with rB15 demonstrating more efficient infection of cells expressing CCR2b ECL3 in the context of CCR5 than did rV1/V2. Therefore the four amino acid differences outwith V1/V2 in the microglia-adapted isolate also affected specific interactions between Env and the coreceptor molecule. Differences in infection efficiency between rBORI and rV1/V2 or rB15 were also found when the N-terminal four amino acids were deleted from CCR5, with rBORI showing reduced infectivity compared with both rV1/V2 and rB15 (Martin et al, 2001). Moreover the microglial-adapted variants showed greatly increased susceptibility to neutralisation by both a panel of human anti-HIV-1 antisera and the CD4-induced monoclonal antibody 17b.

This adaptation to utilise lower levels of CD4 along with changes in the specificity of coreceptor interactions (possibly by alteration of the V1/V2 loop conformation) at the expense of increased susceptibility to antibody neutralisation may be similar to *in vivo* adaptations that occur in HIV-1 populations within the brain. In the periphery, the exposure of neutralisation-sensitive epitopes sometimes associated with adaptation to infect cells with reduced levels of CD4 would probably be lethal to the virus. In contrast, within the relatively immunoprivileged environment of the CNS, concentrations of neutralising antibodies may be sufficiently reduced to allow this adaptation to occur. While other studies have reported that tropism of HIV-1 isolates for microglia and macrophages is similar (Ghorpade et al., 1998a; Hibbitts et al., 1999), adaptations such as these may facilitate infection of both macrophages and microglia in the brain. The phenotypic properties conferred by alterations in N-linked glycosylation sites within V1/V2 are strongly context dependent. Similar patterns of mutations can have different, or even opposite effects depending upon the overall

genetic background of the envelope gene in which they are present (Dumonceaux et al, 1998; Kolchinsky et al, 1999; Labranche et al., 1999; Ly and Stamatatos, 2000; Gorry et al, 2002a).

With regard to evidence for *in vivo* adaptations of HIV-1 within the brain, a neurotropic and neurovirulent isolate has recently been described (Gorry et al, 2001; Gorry et al, 2002a), which was highly fusogenic in MDM and induced neuronal apoptosis in primary brain cultures. Envelope clones from this isolate required only low levels of both CD4 and CCR5 to function in cell-to-cell fusion and single-round infection assays. These Envs also bound CCR5 with high affinity and could do so in the absence of CD4, although they could not facilitate CD4-independent infection. Consistent with the *in vitro* adaptive changes associated with microglial infection described above, this primary virus isolate also was more susceptible to antibody neutralisation (Gorry et al, 2002a). This isolate was also missing a predicted N-linked glycosylation site in the V1/V2 stem of gp120, the loss of which had previously been demonstrated as sufficient to allow CD4-independent infection in the laboratory isolate ADA (Kolchinsky et al., 2001a; Kolchinsky et al, 1999). However in contrast to its effect on ADA Env, reintroduction of this N-linked glycosylation site had no effect on CD4 independent CCR5 binding in the neurotropic primary isolate (Gorry et al, 2002a). This provides further evidence for the unpredictable and strain-specific effects that mutations in HIV-1 Env have upon the conformationally complex interactions between Env and the cellular receptors used for virus entry.

1.10 Molecular Evolution

Molecular evolution comprises the study of rates and patterns of change occurring in genetic material (or the products of this material such as proteins) through evolutionary time, and the mechanisms responsible for these changes. This field of study also encompasses molecular phylogenetics, which involves converting information in sequences into an evolutionary tree for those sequences.

1.10.1 Distance Measures for Nucleotide Sequences

One method for estimating the amount of evolutionary change that has occurred between two nucleotide sequences with a shared common ancestry is to calculate a genetic distance. This is usually measured by the number of nucleotide substitutions between the two sequences (pairwise distance). The simplest distance measure is the p-distance, which corresponds to the proportion of sites at which the two sequences compared are different. A p-distance is therefore obtained by dividing the number of nucleotide differences by the total number of nucleotides compared. Because multiple substitutions can occur at the same nucleotide position, a p-distance may underestimate the actual amount of evolutionary change. Therefore a large number of distance correction methods have been developed, which are designed to more accurately represent the underlying molecular evolutionary processes. The methods used in the investigations described in this thesis were the Jukes Cantor one parameter distance (Jukes and Cantor, 1969) and the Kimura two parameter distance (Kimura, 1980).

The one parameter method of Jukes and Cantor (J-C distances) compensates for multiple substitutions at the same site, and assumes that the rate of nucleotide

substitution is the same for all pairs of the four nucleotides (G, A, T, and C). J-C distances therefore do not take into account differences in the frequencies of transitions (purine to purine and pyrimidine to pyrimidine substitutions), or transversions (purine to pyrimidine or pyrimidine to purine substitutions). This model is suitable for the analysis of HIV sequences that have diverged within a single infected individual, where pairwise distances are relatively small and transitions are unlikely to have saturated to an extent that major underestimates of the rate of sequence change would be made.

The assumption that all nucleotide substitutions occur randomly is unrealistic in many cases, as transitions generally occur with more frequency than transversions. To compensate for this bias the Kimura two parameter model takes into account different rates of transitions and transversions per site (Kimura, 1980).

1.10.2 Molecular Phylogenetics

Evolutionary relationships between sequences are often depicted in the form of a phylogenetic tree. Phylogenetic trees can be rooted or unrooted. In a rooted tree the direction of evolution is indicated, with the root representing the common ancestor of all the remaining sequences. An unrooted tree specifies the relationships among the sequences but does not define an evolutionary path. Most phylogenetic methods produce unrooted trees. However, a root may be added to an unrooted tree by including an outgroup which is evolutionarily related to the sequences under study, but diverged from the other sequences prior to their divergence from one another.

A large number of methods have been described for creating phylogenies from nucleotide sequence data, and there has been extensive debate over which methods

perform best in different circumstances (Page and Holmes, 1998a; Felsenstein, 1988; Felsenstein, 1988). Broadly, methods for creating phylogenies can be divided into character-based methods and distance-based methods. Character-based methods use the individual substitutions among sequences to determine the most likely ancestral relationships. Distance-based methods first convert aligned sequences into a pairwise distance matrix and then input that matrix into a tree building method. Tree building methods can be divided into cluster methods and those that use optimality criteria. Cluster methods follow an algorithm that considers each sequence in turn and adds it to a growing tree (for example unweighted pair-group mean). The second class of tree building methods use optimality criteria to assign a score to all possible trees, and this score is a function of the relationship between tree and data (for example maximum parsimony and maximum likelihood methods). A disadvantage of this class of method is that as increasing numbers of sequences are compared, the total number of possible trees rises rapidly. For example there are 945 possible rooted trees for 6 sequences, over 3×10^7 rooted trees for 10 sequences, and the corresponding figure for 135 sequences is 2.113×10^{267} possible rooted trees. This type of tree building technique is therefore extremely computationally intensive, and becomes unusable with larger data sets.

For the investigations described in this thesis the Neighbour-Joining (N-J) distance-based (clustering) method was used (Saitou and Nei, 1987), which is a simplified version of the minimum evolution method (Saitou and Imanishi, 1989; Rzhetsky and Nei, 1992). The minimum evolution method uses genetic distances that corrects for multiple substitutions at the same site, and a topology showing the smallest value for the sum of all branches is chosen as an estimate of the correct tree. The N-J method

does not compute a value for the sum of all branches; instead the examination of different topologies is embedded in the algorithm. Therefore only one final tree is produced. This method produces an unrooted tree, and usually requires an outgroup to find the root.

1.10.3 Statistical Significance of Tree Topologies

A number of different methods have been proposed for testing the reliability of a tree topology. The most frequently used is the bootstrap test (Felsenstein, 1985). Bootstrapping involves randomly resampling the data on which the tree was based to generate a distribution of data sets (usually several hundred), each of which is used to determine a new tree. During each resampling usually one third of the varied sites are removed and replaced by duplication of others in the same sequence. The frequency that particular branches are observed in the resampled data set then allows probability statements to be attached to the original tree. While there has been extensive debate over the exact significance of bootstrap values, it has been suggested that bootstrap values of 70% or higher may correspond to a 95% confidence level (Hillis and Bull, 1993).

1.11 Aims

HIV-1 is frequently detected in the CNS, where it can cause damage to neural cell types by both direct and indirect mechanisms. Sequence analysis of HIV-1 variants in the brain, usually based upon relatively short regions of the *env* gene, often shows that this subpopulation of virus is distinct from that present in the lymphoid system. Many factors could lead to such divisions and it is not clear which underlies the apparent genetic segregation observed in HIV-1.

Therefore one aim of this project was to further characterise the sequence relationships between HIV-1 in the brain and lymphoid system. In particular, the role of recombination between the two subpopulations was assessed by focussing upon regions of the viral genome that lie both within, and outside the major determinants of HIV cellular tropism (V1/V2 and V3 regions). This analysis included entire HIV-1 envelope genes obtained at limiting dilution. If physical isolation or differences in virus turnover rates are responsible for the separate virus populations observed in the brain, then the same phylogenetic relationships between different variants should be observed regardless of the subgenomic region analysed. However, if lymphoid variants come into regular contact with those in the brain it may be expected that recombinants between the two will arise frequently. These recombinants could then be readily detected by their differing phylogenetic affinities in different subgenomic regions. If this is the case then the genomic areas that maintain a tissue specific grouping are likely to have been selected for, and thus represent a specific adaptation for replication in the brain.

A second important issue regarding HIV sequence relationships in the context of a single infection concerns the manner in which sequences are obtained. Sequence analysis can provide many insights into HIV biology, for example by enhancing our understanding of the viral adaptive processes associated with transmission, immune evasion, development of antiviral resistance, and invasion of new cell and tissue types. As technological advances allow longer sequences to be routinely obtained and larger sequence data sets to be examined, it is likely that our knowledge of each of these processes will be further enhanced. However, there are a number of *in vitro* artefacts associated with polymerase chain reaction (PCR) and molecular cloning

techniques commonly used to obtain HIV sequences. These include the incorporation of point mutations during PCR, recombination during PCR, resampling of clones, and modification of sequences during plasmid expansion in bacteria. Each of these issues has the potential to affect both sequence analysis and interpretations based upon that analysis. Moreover if the sequences are subsequently used in phenotypic assays then the same complications could also arise in these downstream applications. Therefore the second aim of this project was to quantify the effect of each of these *in vitro* artefacts in the context of HIV-1 *env* gene amplification. The results of this analysis demonstrated that all of these factors could lead to considerable alterations in sequences obtained.

Taking into account the sequence alterations associated with PCR amplification and cloning of HIV-1 *env*, and the complex, poorly defined, and often strain dependent adaptations that may underlie HIV-1 neurotropism, it would be useful to have a means of ensuring that *in vitro* artefacts are not present in *env* genes to be used for phenotypic analysis. Therefore the third aim of this project was to develop a method for generating *env* sequences suitable for transfection and expression in mammalian cells that maintained the exact characteristics and genetic linkages present *in vivo*.

2: Materials and methods

2.1 Autopsy samples

Samples were obtained from the Medical Research Council Brain and Tissue Bank of Edinburgh (Western General Hospital, Edinburgh). Frozen samples of lymph node or spleen and several anatomically distinct regions of brain were obtained at autopsies carried out within three days of death. Samples were cut into 1-2 cm³ sections and frozen at -40°C prior to DNA extraction.

2.2 Biological Safety when working with HIV

HIV is classified as a dangerous pathogen of hazard group 3. When working with HIV it is therefore vital to ensure measures are in place that adequately protect laboratory workers and other staff members. As a blood borne virus, infection with HIV may occur via several routes by direct contact with clinical samples containing virus such as blood, semen, vaginal secretions, cerebrospinal fluid, and tissue biopsy material. The most frequent route of infection during laboratory procedures is through cuts from contaminated “sharps” such as syringes, needles, scalpels, and pointed scissors. This allows direct access of virus into the blood or lymphatic system and therefore susceptible cells. Splashes of infectious material onto face and skin can allow viral entry via mucosal membranes or through minor cuts or abrasions. If *in vitro* culture of HIV is undertaken, the viral concentration in the culture supernatant can greatly exceed levels typically found in blood of infected individuals and may therefore pose additional infection risks. Acquisition of HIV infection via ingestion or inhalation has not been documented in a non-laboratory setting, although with high titre virus samples special care must be taken to reduce these potential risks. General safety procedures when handling infectious samples

included the use of a disposable plastic apron, arm cuffs, and two pairs of latex gloves beneath a dedicated lab coat. Autopsy material was dissected using chainmail gloves between two pairs of latex gloves and all sample manipulation was undertaken within a class 1 microbiological safety cabinet. Pseudotype virus production was confined to a class 2 microbiological safety cabinet within a specifically designed category 3 laboratory. Laboratory surfaces were regularly decontaminated with the application of 70% ethanol or 3% Tegador (Tego Hygiene). Contaminated waste material was disposed of by initial chemical treatment where possible with 3% Tegador followed by autoclaving (121°C for 30 minutes at 3 atmospheres) then incineration.

2.3 DNA Extraction

0.5-1 cm³ blocks were dissected from frozen autopsy material and added to 1.5 ml eppendorf tubes containing 500 µl lysis buffer (0.11 M NaCl₂, 55 mM Tris pH 8.0, 1.1 mM ethylenediamine tetraacetic acid [EDTA] pH 8.0, 0.55% sodium dodecyl sulphate [SDS], 1 mg/ml Proteinase K, 40 µg/ml poly A), mixed well by vortexing, and then incubated at 65°C for two hours or until the material was completely dissolved. After lysis, 450 µl of water saturated phenol (Sigma) was added to the extraction eppendorf and vortexed for 5 minutes. The tube was then centrifuged at room temperature for 10 minutes at 12000g. The upper aqueous layer was transferred to a clean eppendorf tube and phenol extraction repeated as above. The aqueous layer was transferred to a fresh eppendorf tube containing 450 µl chloroform/iso-amylalcohol (Rathburn) 50:1, mixed by vortexing for 2 minutes and centrifuged at room temperature for 10 minutes at 12000g. The aqueous layer was then transferred

to a fresh eppendorf tube containing 40 μ l 3 M sodium acetate (NaAc) pH 5.2 and 800 μ l cold (-20°C) 100% ethanol (BDH). This tube was mixed by inversion and then left overnight at -20°C to allow precipitation of nucleic acid. After precipitation the tube was centrifuged at 0°C for 10 minutes at 12000g and the supernatant discarded. The pellet was then washed with 80% (v/v) ethanol and centrifuged at 0°C for 10 minutes at 12000g. The supernatant was discarded and the pellet dried at 37°C for 5 to 10 minutes. The DNA pellet was re-suspended in 50 μ l of sterile (DNAase/RNAase free) water and left at room temperature for at least 10 minutes before use to ensure adequate solubilisation of the nucleic acid.

2.4 Quantitation of Extracted DNA

The concentration and approximate purity of DNA in each sample was quantified using spectrophotometric absorbance readings at wavelengths of 260 nm and 280 nm. Samples were diluted 1 in 200 with sterile water and absorbance measured in a spectrophotometer (GeneQuant II, Pharmacia Biotech). The concentration of DNA was calculated from the equation: $A_{260} \times D \times 50 = \text{DNA concentration } (\mu\text{g/ml})$, where A_{260} = OD values at 260 nm, D is the dilution factor (200), and 50 is equivalent to the concentration ($\mu\text{g/ml}$) of double-stranded DNA at A_{260} of 1.0. The ratio of optical density (OD) at 260 nm to 280 nm gave an indication of DNA purity in each of the extracted samples. Pure DNA preparations have an A_{260}/A_{280} ratio of approximately 1.8.

2.5 Polymerase Chain Reaction

2.5.1 General procedures

The nested polymerase chain reaction is an extremely specific and sensitive molecular technique allowing the amplification of a single target molecule within a complex mixture of DNA sequences. While ideal for a wide range of biological applications the high sensitivity of this procedure gives rise to an inherent risk of sample contamination by exogenous DNA during laboratory procedures. To minimise this risk all buffers for PCR reactions were prepared in a dedicated room where no nucleic acid extractions were performed. Primary and secondary PCR reactions and the cloning and subsequent amplification of PCR products were all undertaken in separate rooms with dedicated lab coats. Latex gloves were changed frequently throughout procedures. Known positive and negative DNA controls and an additional negative control containing no DNA were included in every set of reactions to provide a comprehensive test for contamination of samples or buffers.

All thermal cycling reactions were performed on a Genius PCR machine (Techne). Where nested PCR was performed, 2 μ l of primary PCR product was transferred for amplification with second round primers and the same reaction conditions were used for primary and secondary PCRs. All reactions were overlaid with 1 drop of mineral oil prior to thermal cycling to prevent evaporation of reaction mixtures.

2.5.2 Primer sequences

General considerations when designing oligonucleotide primers for PCR include primer length (generally between 15 and 30 bases), close sequence identity with the template DNA sequence (especially when heterogeneous DNA templates are being

amplified), and avoidance of mismatches with template DNA at the 3' base position. If possible similar annealing temperatures for forward and reverse primers, 40-60% G + C content, avoidance of runs of Gs or Cs (to reduce internal secondary structure), and avoidance of complementary sequences between primers (to avoid primer-dimer formation) is also desirable. The sequences of the oligonucleotide primers used in this study are shown (Table 2.1) along with the 5' base position of the primer-binding site on the relevant template molecule.

Code	Primer Sequence (5' to 3')	Binding site
531	GCGAGAGCGTCAGTATTAAGCGG	1012: p17 ^{gag} 1
532	GGGAAAAAATTCGGTTAAGGCC	1052: p17 ^{gag}
533	CTTCTACTACTTTTACCCATGC	1528: p17 ^{gag}
534	TCTGATAATGCTGAAAACATGGG	1577: p17 ^{gag}
401	GAGGATATAATCAGTTTATGG	7372: V1/V2
402	CATCAAAGCCTAAAGCCATG	7393: V1/V2
403	CAATAATGTATGGGAATTGG	7883: V1/V2
333	AATGTACTGTGCTGACATT	7969: V1/V2
332	TACAATGTACACATGGAATT	7964: V3
306	TGGCAGTCTAGCAGAAGAAG	8019: V3
307	CTGGGTCCCCTCCTGAGG	8422: V3
308	ATTACAGTAGAAAAATCCCC	8475: V3
G	AGATCTTCAGACTTGGAGGAGG	8798: gp41
I	GAGTAAGTCTCTCAAGCGGTGG	9795: gp41
ENV A	GGCTTAGGCATCTCCTATGGCAGGAAGAA	6622: env
ENV B	AGAAAGAGCAGAAGACAGTGGCAATGA	6972: env
ENV M	TAGCCCTTCCAGTCCCCCTTTTCTTTTA	10473: env
ENV N	CTGCCAATCAGGGAAGTAGCCTTGTGT	10554: env
N1	AGCAGTAGCTGAGGGRACAGATAG	9978: nef
NEF F	TGCTRTAAGATGGGTGGCAAGTGG	10082: nef
NEF F2	TAAGATGGGTGGCAAGTGGTCAA	10087: nef
N4	CTGGCCCTGGTGTGTARTTCTGCC	10576: nef
NEF R	TGGAAAGTCCCCAGCGGAAAGTCC	10989: nef
NEF RO	GCTGCTTATATGCAGGATCTGAGG	11099: nef
GFP1	GTGGATAACCGTATTACCGCCATGC	4705: pEGFP-C1 ²
GFP2	CCGCTAGCGCTACCGGTCGCCACC	589: pEGFP-C1
GFP3	CTA CTTGTACAGCTCGTCCATGCCG ³	1332: pEGFP-C1
GFP4	GGATTTTGCCGATTTCCGGCTATTG	1739: pEGFP-C1

LUC F	GGTAAAGCCACCATGGAAGACGCC	268: luciferase ⁴
LUC F2	CCATGGAAGACGCCAAAAACATAA	278: luciferase
LUC R2	GCGATCTTTCCGCCCTTCTTGCC	1925: luciferase
LUC R	GAATTACACGGCGATCTTTCCGCC	1935: luciferase
3.1FO	GATTTTGCCGATTTCCGGCTATTG	4926: pCR3.1 ⁵
3.1F	GCGCGCGTTGACATTGATTATTGAC	1: pCR3.1
3.1R	CTGCGTTATCCCCTGATTCTGTGG	1067: pCR3.1
3.1RO	TTACGGTTCCTGGCCTTTTGCTGG	1113: pCR3.1
T7	TAATACGACTCACTATAGGG	638: pCR3.1
BGH	TAGAAGGCACAGTCGAGG	790: pCR3.1

Table 2.1. Oligonucleotide primer sequences and 5' nucleotide binding sites on template DNA.

¹HIV sequence positions correspond to the HXB2 genome on the Los Alamos National Laboratory HIV Database. ²pEGFP-C1, genbank accession U55763. ³Shaded bases represent alteration of vector sequence to incorporate a stop codon. ⁴Luciferase gene in pGL3-control vector, genbank accession U47298. ⁵pCR3.1-Uni mammalian expression vector (numbering corresponds to Invitrogen literature).

2.5.3 PCR of p17^{gag}, V1/V2, V3, nef and EGFP

The enhanced green fluorescent protein (EGFP) gene (both with and without flanking promoter and poly [A] sequences) was amplified using 2 µl of a 10⁻³ dilution (approximately 10 ng) of pEGFP-C1 vector (Clontech) as template for PCR with a single set of primers. HIV subgenomic regions were amplified by nested PCR using template DNA extracted from autopsy samples.

Reagents

Taq polymerase in storage buffer B (Promega) – 5 Units/µl enzyme, 20 mM Tris-HCl (pH8), 100mM KCl, 0.1 mM EDTA, 1 mM DDT, 50% glycerol, 0.5% Nonidet-P40, 0.5% Tween 20

10 × PCR reaction buffer with MgCl₂ (Promega) – 100 mM Tris-HCl (pH 9.0), 500 mM KCl, 15 mM MgCl₂, 1% Triton X-100

Protocol

Reaction mixtures contained 5 µl of 10 × reaction buffer, 30 µM of each dNTP (dGTP, dATP, dTTP, and dCTP), 1 Unit (U) of *Taq* polymerase, 0.25 µM of each primer and between 1 and 5 µl of sample DNA. All PCR mixes were set up in 50 µl reactions with the final volume made up with nuclease free water. Primers used to amplify primary and secondary reaction products are shown (Table 2.2).

Thermal cycling conditions

p17^{gag} and *nef* regions

94°C for 18 sec, 55°C for 21 sec, 72°C for 1.5 min × 30 cycles followed by 72°C for 10 min × 1 cycle

V1/V2 and V3 regions:

94°C for 18 sec, 50°C for 21 sec, 72°C for 1.5 min × 25 cycles followed by 72°C for 10 min × 1 cycle

EGFP:

94°C for 45 sec, 55°C for 45 sec, 72°C for 3 min × 25 cycles followed by 72°C for 10 min × 1 cycle

Sequence Amplified	Primary PCR		Secondary PCR	
	Forward	Reverse	Forward	Reverse
p17 ^{gag}	531	534	532	533
V1/V2	401	333	402	403
V3	332	308	306	307
<i>env</i>	ENV A	ENV N	ENV B	ENV M
<i>nef</i>	N1	NEFRO	NEFF	NEFR
Promoter-EGFP-poly (A)	GFP1	GFP4	-	-
EGFP	GFP2	GFP3	-	-
Luciferase	LUCF	LUCR	-	-

Table 2.2. Primers pairs used in PCR and sequencing reactions.

2.5.4 PCR of full-length *env* and Luciferase genes

PCR of the *env* gene was performed using template DNA extracted from autopsy samples or control clones. Control clones used were CAM-1 (full-length molecular clone of the Cambridge isolate also referred to as C-HIV-1) and LAI (full-length 12 Kb insert of HIV-1 LAI) kindly provided by Dr Karpas and Dr Becket respectively via the NIBSC Centralised Facility for AIDS Reagents supported by EU Programme (contract BMH4 97/2515) and the UK Medical Research Council. Plasmid DNA was extracted as in section 2.8.5 and quantified by V3 limiting dilution PCR (see 2.5.3). Several aliquots were made whereby either a single genotype or an equimolar mix of the two genotypes (approximately 3×10^4 copies of each) was added to 1 μ g of HIV negative peripheral blood mononuclear cell (PBMC) DNA. Various dilutions of these mixes were then used as templates for *env* PCR reactions (as described in Chapter 4). The same reaction conditions were used in a single round PCR reaction to amplify the firefly luciferase gene (*luc*⁺) using 2 μ l of a 10^{-5} dilution (approximately 100 pg) of pGL3-control vector (Promega) as template.

Reagents

10 × PCR reaction buffer (with MgCl₂) (Roche)

Expand High Fidelity Polymerase (Roche) – 3.5 U/μl enzyme, 20 mM Tris-HCl (pH 7.5), 100mM KCl, 1 mM DTT, 0.1 mM EDTA, 0.5% Tween 20, 0.5% Nonidet-P40, 50% glycerol

Protocol

Reaction mixtures contained 5 μl of 10 × reaction buffer with MgCl₂, 60 μM of each dNTP, 2 units of Expand High Fidelity Polymerase, 0.5 μM of each primer and between 1 and 5 μl of sample DNA with the final volume of 50 μl made up with nuclease free water.

Thermal cycling conditions

94°C for 45 sec, 50°C for 45 sec, 72°C for 4 min × 20 cycles, then 94°C for 45 sec, 50°C for 45 sec, 72 °C for 4min15 sec × 15 cycles, followed by 72°C for 10 min × 1 cycle

2.6 Agarose gel electrophoresis

Reagents

10 × TBE – 108g Tris base (BDH), 55g boric acid (Molecular Biology certified; Kodak), 40ml 0.5M EDTA (Molecular Biology certified; Kodak). Made up to 1 litre with distilled water

Agarose gel – Agarose (Sigma), made with 1 × TBE containing ethidium bromide (0.1 μg/ml)

6 × loading buffer – 15% Ficoll 400, 0.03% bromophenol blue, 0.03% xylene cyanol, 0.4% orange G, 10 mM Tris-HCl (pH 7.5), 50 mM EDTA

Protocol

DNA was fractionated according to size in a 1-2% agarose gel using flatbed apparatus and $1 \times$ TBE as electrophoresis buffer. Gels were run at 150V for 20 minutes to 2 hours depending on the expected size of the DNA. Ethidium bromide was included in the gel to act as an intercalating agent, binding to DNA and fluorescing under short wave (254 nm) ultra violet light therefore allowing PCR products, restriction digests, and clones to be visualised as discrete bands. DNA samples were suspended in a final concentration of $1 \times$ loading buffer prior to electrophoresis and 100 bp or 1 Kb DNA ladders (Promega) were run beside samples to assess the length of products.

2.7 Quantitation of HIV-1 proviral DNA

HIV-1 DNA was initially semi quantified by using nested PCR on serial 10-fold dilutions of extracted DNA with V3 primers known to be capable of detecting a single proviral copy. In some cases this was followed by the more accurate limiting dilution PCR method (Simmonds et al., 1990). Ten PCR replicates were performed at the last positive dilution and the proviral load was estimated assuming a Poisson dilution by the formula $-\ln(1-p)/d$ (where p = proportion of positive samples and d = dilution). The number of proviral copies per million cells was calculated on the basis that a human diploid cell contains 6.6 pg DNA.

2.8 Cloning of PCR products

The cloning of PCR products was performed using a unidirectional eukaryotic TA cloning kit (Invitrogen). This kit contains pCR3.1, a vector allowing TA cloning of

complete open reading frame PCR products in the correct orientation for expression when the resultant clones are transfected into eukaryotic cells. The left hand arm of pCR3.1 does not have a 5' phosphate group and in order for ligation to occur the PCR products must be amplified with forward primers that have a 5' phosphate group chemically or enzymatically added, therefore assuring correct (forward) orientation of insert.

2.8.1 Phosphorylation of PCR products

5' phosphates were added to forward primers either chemically during primer synthesis (Oswel) or enzymatically using T4 polynucleotide kinase. The enzymatic phosphorylation reaction contained 50-100 pmoles of forward primer, 1 μ l of 10 \times kinase buffer (500 mM Tris-HCl (pH 7.5), 100 mM MgCl₂, 50 mM dithiothreitol, 1 mM spermidine), 1 μ l 10 mM ATP, and sterile water to make up to 9 μ l total volume. One μ l of T4 polynucleotide kinase (10 U/ μ l) was added and the reaction mixed gently then incubated at 37°C for 30-40 minutes. The reaction was then incubated at 94°C for 5 minutes to inactivate the kinase and prevent phosphorylation of the reverse primer during PCR. Enzymatically phosphorylated primers were used for PCR immediately after production.

2.8.2 Ligation reaction

The amount of PCR product needed to ligate with 60 ng (20 fmoles) of pCR3.1 was calculated using the formula **X ng PCR product = 2 (Y bp PCR product) (60 ng pCR3.1 vector) / (size in bp of the pCR3.1 vector: 5020)** where "X" ng is the amount of PCR product of "Y" bp to be ligated for a 1:2 (vector:insert) molar ratio.

The ligation reaction was composed as follows:

Fresh PCR product (Less than 1 day old)	X μ l
10 \times Ligation Buffer – 60 mM Tris-HCl (pH 7.5), 60 mM MgCl ₂ , 50 mM NaCl, 1.0 mg/ml bovine serum albumin, 70 mM β -mercaptoethanol, 1 mM ATP, 20 mM dithiothreitol, 10 mM spermidine	1 μ l
PCR3.1-Uni vector (30 ng/ μ l)	2 μ l
Sterile water	To a final volume of 9 μ l
T4 DNA ligase (4 units/ μ l)	1 μ l

Table 2.3. Composition of ligation reaction.

All ligation reactions were incubated overnight at 15°C then used immediately or stored at -20°C.

2.8.3 Transformation

Reagents

SOC Medium – 2% Tryptone, 0.5% yeast extract, 0.005% NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM glucose

LB (Luria-Bertani) Medium – 1% Tryptone (Difco), 0.5% Yeast extract (Gibco), 1% NaCl, made up to 1 litre with distilled water

LB agar plates - 1% Tryptone, 0.5% Yeast extract, 1% NaCl, 1.5% agar, made up to 1 litre with distilled water

One Shot TOP10F' competent cells – genotype: F' {*lacI*^q Tn10 (Tet^R)} *mcrA* Δ (*mrr*-*hsdRMS*-*mrcBC*) ϕ 80*lacZ* Δ M15 Δ *lacX74* *deoR* *recA1* *araD139* Δ (*ara-leu*)7697 *galU* *galK* *rpsL* *endA1* *nupG*

Protocol

One vial containing 50 μ l of competent cells was thawed on ice for each transformation reaction. Two microlitres of ligation reaction was added to the competent cells and the mixture gently stirred before incubation for 30 minutes on ice. The vials were then heated for exactly 30 seconds at 42°C and immediately placed back on ice. Two hundred and fifty μ l of room temperature SOC media was added to the tubes and the vials were incubated at 37°C for 1 hour at 225 rpm in a shaking incubator. Between 10 and 50 μ l of the transformed cells were then plated onto LB agar plates containing 50 μ g/ml ampicillin and left at room temperature for 20 minutes. The plates were then inverted and incubated at 37°C overnight before storage at 4°C.

2.8.4 Colony Screening

Bacterial colonies transformed with V1/V2 ligations were PCR screened for the presence of insert using primers 402 and 403. Colony screening for the presence of *env* gene clones required two separate PCR reactions. The first reaction included primers Env B and 403, therefore verifying the presence of the 5' end of the gene. The second reaction included primers N2 and ENV M, indicating the 3' end of the gene was also present. A small proportion of a bacterial colony was picked with a sterile toothpick and added directly into PCR reaction mixtures for amplification. PCR reaction mixture composition (with the exception of primers) and thermal cycling parameters were as for amplification of V3 (Section 2.5.3).

2.8.5 Plasmid DNA Mini-preps

Bacterial cultures for plasmid mini-preps were prepared by inoculating 2-3 ml of LB media containing 50 µg/ml of ampicillin with a single colony in a 15 ml sterile tube to ensure adequate aeration for growth. The culture was incubated at 37°C with shaking for 12-16 hours. A plasmid mini prep kit (Hybaid) was then used to extract the plasmid DNA from the bacteria as follows:

Reagents

Pre-Lysis buffer – (50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 100 µg/ml RNAase A

Alkaline Lysis Solution – 200 mM NaOH, 1% SDS (w/v)

Neutralising Solution – Containing acetate and guanidine hydrochloride

Binding Buffer – Containing silica gel matrix

Wash Solution – Containing ethanol, NaCl, EDTA, and Tris-HCl

Protocol

One ml of bacterial culture was transferred to an eppendorf and centrifuged at 6000g for 5 minutes. The supernatant was completely removed and discarded then the bacterial pellet was resuspended in 50 µl pre-lysis buffer. To lyse the cells 100 µl of alkaline lysis solution was added and mixed thoroughly into the cell suspension. A total of 75 µl neutralising solution was added and the tube vortexed to ensure complete mixing of contents. The tube was then centrifuged for 2 minutes at 12000g to pellet the white precipitate and the supernatant removed and transferred to a spin column. Binding buffer was shaken before use to resuspend the silica gel matrix and 250 µl of this buffer was added to the spin filter and mixed with the plasmid DNA

supernatant. The spin tube was centrifuged for 1 minute at 12000g to collect the liquid at the bottom of the vial. Three hundred and fifty μ l of wash solution was added and the spin filter was centrifuged again for 1 minute. In order to dry the pellet, the liquid in the collection vial was discarded and the spin filter was centrifuged for 1 minute at 12000g. The spin filter was then transferred to a new catch tube and 50 μ l of sterile water was added to the membrane. The tube was sealed, vortexed briefly, and centrifuged for 30 seconds to collect plasmid DNA in the catch tube.

2.8.6 Plasmid DNA Midi-preps

For isolation of up to 100 μ g of plasmid DNA larger bacterial culture volumes were required. A culture of 2-3 ml LB media containing 50 μ g/ml ampicillin was inoculated with a single colony and cultured in a sterile 15 ml tube for approximately 8 hours at 37°C with vigorous shaking (>200 rpm). This starter culture was diluted 1/500-1/1000 and used to inoculate 25 ml of LB medium with 50 μ g/ml ampicillin in a sterile container of at least 100 ml total volume. The second culture was grown at 37°C for 12-16 hours with vigorous shaking. A QIAfilter plasmid midi kit (Qiagen) was used to extract plasmid DNA as follows:

Protocol

The bacterial cells were harvested by centrifugation at 6000g for 15 min at 4°C after which the supernatant was discarded and the pellet resuspended in 4 ml buffer P1. Four ml of buffer P2 was then added to lyse the cells and the lysate mixed by inverting the tube 4-6 times before incubation for 5 minutes at room temperature. Four ml of chilled buffer P3 was added to the lysate and the tube inverted 4-6 times

to precipitate genomic DNA, proteins, and cell debris. The lysate was then poured into a QIAfilter cartridge and incubated at room temperature for 10 minutes. A QIAGEN-tip 100 was equilibrated by applying 4 ml of buffer QBT and allowing the column to empty by gravity flow. The cell lysate was filtered through the QIAfilter cartridge into the equilibrated QIAGEN-tip 100 and allowed to enter the resin by gravity flow. The tip was washed twice with 10 ml of buffer QC then the plasmid DNA eluted with 5 ml of buffer QF. DNA was then precipitated by mixing 3.5 ml room temperature isopropanol with the eluted DNA and centrifuging at 15000g for 30 minutes at 4°C. The supernatant was decanted off and the DNA pellet washed with 2 ml room temperature 70% ethanol before centrifugation at 15000g for 10 minutes. The ethanol was then removed and the pellet air-dried for 10 minutes before redissolving the DNA in 100 µl sterile water.

2.9 Restriction digests

2.9.1 *Pme* I digest

pCR3.1 contains two *Pme* I sites, one on each side of the insert. This restriction enzyme was used to cut the vector and verify that clones contain inserts of an appropriate size.

Reagents

Pme I restriction enzyme (New England Biolabs) – 10 U/µl *Pme* I enzyme in storage buffer (100 mM NaCl, 10 mM Tris-HCl (pH7.4), 0.1 mM EDTA, 1 mM dithiothreitol (DTT), 200 µg/ml BSA, and 50% glycerol)

10 × New England Biolabs buffer 4 – 50 mM potassium acetate, 20 mM Tris acetate, 10 mM magnesium acetate, 1 mM DTT

Protocol

2 µl cloned DNA was mixed with 2 units of *Pme* I, 1 µl 10 × reaction buffer 4, 0.1 µg/ml final concentration of Bovine serum albumin (BSA), and the final volume of 10 µl made up with sterile water. The reaction was incubated for 4 hours at 37°C and run on an agarose/ethidium bromide gel for visualisation of products.

2.9.2 *Hind* III and *Xmn* I single and double digests:

Full-length *env* PCR products were cleaved with *Hind* III, *Xmn* I, or a combination of both restriction enzymes.

Reagents

Hind III restriction enzyme (Promega) – 10 U/µl *Hind* III enzyme in storage buffer (10 mM Tris-HCl, (pH7.4), 250 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 0.5 mg/ml BSA and 50% glycerol)

Xmn I restriction enzyme (Promega) – 10 U/µl *Xmn* I enzyme in storage buffer (10 mM Tris-HCl (pH7.5), 100 mM NaGlutamate, 0.1 mM EDTA, 1 mM DTT, 0.5 mg/ml BSA, and 50% glycerol)

10 × reaction buffer B (Promega) – 60 mM Tris-HCl (pH7.5), 500 mM NaCl₂, 10 mM DTT

Protocol

Single and double digests were performed in 30 µl total volumes. Digests were set up with 8 µl of secondary PCR product (approximately 800 ng DNA), 3 µl of 10 ×

reaction buffer B, 2 units of (each) enzyme, 0.1 µg/µl concentration of BSA with the final reaction volume made up with sterile water. The reaction was incubated at 37°C for 4 hours and products visualised under UV light after electrophoresis on an agarose/ethidium bromide gel.

2.10 Overlap Extension PCR reactions

Linear expression constructs (LECs) were produced by overlap extension PCR reactions. Detailed diagrams of vectors used and PCR primer binding sites are shown in chapter 5.

2.11 PCR using ligation reactions as template

pCR3.1 ligation reactions were set up as described in section 2.8.2 for EGFP, *luc*⁺, *nef*, and *env* genes amplified by PCR. In each case 2 µl of a 10⁻³ dilution of the ligation reaction products was used as a template for further PCR reactions. Two separate overlapping single round PCR reactions (PCR A and PCR B) were set up with each ligation product. The PCR conditions were the same as those for the *env* gene amplification but *Taq* polymerase and the appropriate 10 × buffer replaced Expand High Fidelity Polymerase and buffer. The primers used for amplification of these products are shown in Table 2.4.

Template Ligation	Primers for PCR A	Primers for PCR B
EGFP	3.1FO and GFP3	GFP2 and 3.1RO
Luciferase	3.1FO and LUCR2	LUCF2 and 3.1RO
nef	3.1FO and N4	NEFF2 and 3.1RO
env	3.1FO and 332	308 and 3.1RO

Table 2.4. Primer pairs used in construction of linear expression constructs

2.12 Extension reactions

Extension reactions were performed for EGFP and Luciferase by mixing 5 μ l of PCR product A with 25 μ l of product B, 1.5 nmol of each dNTP and 1 unit of *Taq* polymerase. *nef* and *env* extensions were composed by adding 5 μ l PCR product A to 5 μ l product B, 5 μ l 10 \times reaction buffer B, 3 nmol of each NTP, and 2 units of *Taq* polymerase with the final reaction volume of 60 μ l made up with sterile water. Extension mixtures were then cycled at 94°C for 24 sec, 72°C for 4 min \times 10 cycles followed by 72°C for 10 min \times 1 cycle.

2.13 Final Amplification

Extension products were subjected to a final amplification reaction using primers 3.1F and 3.1R. For EGFP and luciferase genes 1 μ l of extension reaction was taken for further amplification, for *nef* and *env* genes 10 μ l of extension reaction was carried over. Extended products were added to 5 μ l 10 \times reaction buffer B, 1.5 nmol of each dNTP, 10 pmol of each primer and 1 unit of *Taq* polymerase with 50 μ l final

volume made up with sterile water. The mixtures were overlaid with 1 drop of mineral oil and thermal cycled at 94°C for 45 sec, 55°C for 45 sec, and 72°C for 4 min 15 sec × 15 cycles followed by 72°C for 10 min × 1 cycle.

2.14 PCR product purification

Prior to transfection into eukaryotic cells, linear expression constructs were purified using a QIAquick PCR purification kit (Qiagen). This procedure involves binding the DNA to a silica-gel membrane within a spin column. After binding to the membrane the DNA is washed to remove impurities such as salts, enzymes, unincorporated nucleotides, oils, and detergents that may interfere with transfection procedures. The DNA can also be concentrated during the purification process by adding the contents of several PCR reactions onto a single membrane before elution. In this case three 50 µl amplification reaction products were added to a single column for elution in 50 µl of sterile water.

Protocol

5 volumes of buffer PB was mixed with 1 volume of PCR sample and the mixture added to a QIAquick spin column. The column was then centrifuged at 12000g for 30-60 seconds and the flow-through discarded. 0.75 ml of wash buffer PE was added to the column and it was centrifuged at 12000g for 30-60 seconds. The flow-through was discarded and residual wash buffer removed by centrifuging the column again at 12000g for 30-60 seconds. The column was then placed in a clean tube and 50 µl of sterile water was added to the membrane. Finally the tube was centrifuged at 12000g for 60 seconds to collect eluted DNA.

2.15 DNA sequencing

2.15.1 Manual Sequencing

Manual DNA sequencing was performed using a thermo sequenase radiolabeled terminator cycle sequencing kit (USB). This kit is based upon the chain termination method of Sanger (Sanger et al., 1977) using cycle sequencing (Murray, 1989). The procedure uses a thermostable polymerase to extend a single primer along a DNA template by thermal cycling with the inclusion of ^{33}P radiolabeled 2',3'-dideoxynucleoside-5'triphosphate nucleotide analogues (ddNTPs) as chain terminators. This results in both the radioactive labelling of the extended DNA strand, and the production of different lengths of DNA. Sequencing products can then be separated by electrophoresis on a polyacrylamide gel. After running, the gel is incubated beside photographic film and bands visible upon development of this film correspond to the presence of the same base in the target DNA sequence as the ddNTP added to that reaction. Thus when the four separate sequencing reactions corresponding to each base (ddGTP, ddATP, ddTTP, and ddCTP) are run side by side, the original sequence of the DNA template can be read.

Reagents

Thermo sequenase DNA polymerase – 4 U/ μl thermo sequenase DNA polymerase enzyme, 50 mM Tris-HCl (pH 8.0), 1 mM DTT, 0.1 mM EDTA, 0.5% Tween-20, 0.5% Nonidet P-40, 50% glycerol

Reaction buffer (concentrate) – 260 mM Tris-HCl (pH9.5), 65 mM MgCl_2

dGTP nucleotide master mix – 7.5 μM dGTP, 7.5 μM dATP, 7.5 μM dTTP, 7.5 μM dCTP

ddGTP – 0.3 μM ($\alpha\text{-}^{33}\text{P}$) ddGTP

ddATP – 0.3 μM ($\alpha\text{-}^{33}\text{P}$) ddATP

ddTTP – 0.3 μM ($\alpha\text{-}^{33}\text{P}$) ddTTP

ddCTP – 0.3 μM ($\alpha\text{-}^{33}\text{P}$) ddCTP

Stop solution – 95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF

Shrimp alkaline phosphatase (SAP, Amersham) – 1U/ μl enzyme, 25 mM Tris-HCl (pH 7.5), 1 mM MgCl_2 , 0.1 mM ZnCl_2 , 50% glycerol

Exonuclease I (Amersham) – 10 U/ μl enzyme, 20 mM Tris-HCl (pH 7.5), 0.5 mM EDTA, 5 mM 2-mercaptoethanol, 50% glycerol

Protocol

For sequencing of PCR products the product was first treated with a combination of SAP and Exonuclease I to eliminate any primer or dNTPs, which were not incorporated into the PCR product. One μl of SAP and 1 μl of exonuclease I was added to 5 μl of PCR and the mixture incubated at 37°C for 15 minutes followed by 80°C for 15 minutes. Termination mixes were made by adding 2 μl of dGTP master mix to 0.5 μl of each ddNTP (ddGTP, ddATP, ddTTP, and ddCTP – one for each per sequence). Four tubes were labelled ("G", "A", "T", and "C") and 2.5 μl of each termination mix was added. Reaction mixture composed of 0.5 μl reaction buffer, 0.75 μl DNA, 10 pmoles primer, 0.5 μl thermo sequenase enzyme, and 3.1 μl sterile water was added to each termination tube, mixed well and covered with 10 μl of

mineral oil. The reactions were subjected to 30 cycles at 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 1 minute. After thermal cycling 6 µl of each reaction was transferred to a fresh tube containing 4 µl of stop solution.

2.15.2 Acrylamide Gel Electrophoresis

Reagents

10 × Sanger TBE – 324 g Tris base, 85 g boric acid, 19 g EDTA, made up to 2 litres with distilled water

6% acrylamide gel – 21 g Urea (BDH), 6 ml Sequagel XR concentrate (National Diagnostics), 5 ml 10 × Sanger TBE, 0.05 g ammonium persulphate (APS) (Sigma), made up to 50 ml with distilled water then 20µl N,N,N',N' Tetramthylethylenediamine (TEMED, Sigma) added

Protocol

The sequencing reactions were run on a 6% acrylamide gel. Glass plates were cleaned with methanol then acetone and a pair of flat spacers were used to assemble the plates. The gel mixture was poured carefully to avoid producing bubbles and left to polymerise for at least 2 hours. The gel was pre run at 75 volts for 10 minutes using 1 × Sanger TBE as the electrophoresis buffer. Samples were heated at 90°C for 5 minutes and loaded on the gel, which was run at 75 volts for 0.5-1.5 hours. After electrophoresis the gel was disassembled and dried before exposure to X-ray film (Kodak) for 1-2 days depending on the intensity of radioactive signal.

2.15.3 Automated Sequencing

Automated sequencing reactions were performed using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) and run on an ABI Prism 3100 genetic analyser (Perkin Elmer ABI biosystems). This system follows the same principle as manual cycle sequencing. However, each termination nucleotide analogue is labelled with a different fluorescent marker, allowing sequencing reactions to be performed in a single tube. Primers used for sequencing of *env* gene and subgenomic regions are shown in figure 2.1.

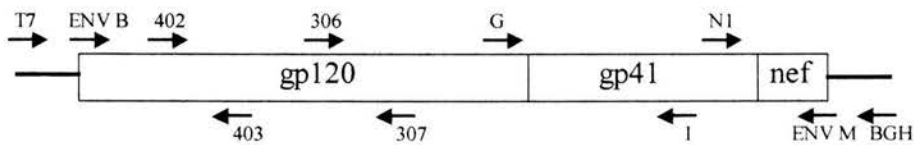


Figure 2.1. Schematic representation of primer binding sites on *env* gene. Primers used to sequence *env* PCR products. For sequencing of clones T7 and BGH were used instead of ENV B and ENV M respectively.

Reagents

Terminator Ready Reaction Mix – A-Dye Terminator labelled with dichloro[R6G], C-Dye Terminator labelled with dichloro[ROX], G-Dye Terminator labelled with dichloro[R110], T-Dye Terminator labelled with dichloro[TAMRA], deoxynucleoside triphosphates (dATP, dCTP, dITP, dUTP), AmpliTaq DNA polymerase, MgCl₂, Tris-HCl buffer (pH 9.0)

Protocol

Reactions were composed of 4 µl of Terminator Ready Reaction Mix, 50-500 ng DNA, 1.6 pmol of primer, and the final reaction volume of 10 µl made up with

sterile water. The mixture was overlaid with 1 drop of mineral oil and subjected to 25 × cycles of 96°C for 30 sec, 50°C for 15 sec, and 60°C for 4 minutes. After thermal cycling the products were transferred to a clean tube and ethanol precipitated by addition of 25 µl cold (-20°C) 100% ethanol and 1 µl of sodium acetate (pH 4.5). The contents were mixed and the tube incubated for 3-16 hours at -20°C before centrifugation at 12000g for 30 minutes at 4°C. The supernatant was removed and the DNA pellet washed by addition of 125 µl of cold 70% ethanol then centrifuged at 12000g for 10 minutes at 4°C. Ethanol was then removed completely and the pellet allowed to air dry before being sent for automated sequencing.

2.16 Sequence Analysis

Sequencing gels were read and manually aligned using the Simmonic 2000 computer package developed by Professor Peter Simmonds. Automated sequencing files were first manually edited using the Chromas ABI sequence editor package (version 1.42) before importing into Simmonic 2000.

Phylogenetic trees were created from the aligned sequences using the Molecular Evolutionary Genetic Analysis (MEGA) computer package (Versions 1.1 and 2.1). A genetic distance based method was used to create data sets for the construction of neighbour-joining phylogenetic trees. Because of the small number of possible nucleotides at any given position in a DNA sequence (four: G, A, T, and C) there is the problem that evolutionary convergence into the same state by two unrelated sequences occurs relatively frequently. The result of this is a significant underestimation of genetic distances, a bias that increases as the sequences compared become more divergent. It is therefore not sufficient to adopt a measure based

entirely on the observed similarity between strains and it is essential to estimate the real divergence from the data by using an estimator that corrects for multiple substitutions at each nucleotide site. In the case of HIV sequences that have diverged during a single infection there are comparatively few nucleotide substitutions and the relatively simple "one parameter" model of Jukes and Cantor is sufficient. In this model it is assumed that any nucleotide is equally likely to be substituted by any other and the estimated number of nucleotide substitutions (K) that have occurred since the two sequences diverged is given by the formula $K = -3/4 \ln (1-4/3\lambda)$ where λ is the observed fraction of sites at which the two sequences differ.

Phylogenetic inferences were made by the construction of a Neighbour-Joining (N-J) tree from the distance matrix. This method is a simplified version of the minimum evolution method (Saitou and Imanishi, 1989; Rzhetsky and Nei, 1992) and uses a complex algorithm to construct a tree from the distance matrix. Since this method produces an unrooted tree, orientation is provided by including an outgroup in the form of an epidemiologically unlinked sequence that is equally divergent from all of the sequences included for analysis. The robustness of the observed groupings is determined by a bootstrap resampling procedure. This method has the advantage over more traditional means of determining variability in that no theoretical models or mathematical assumptions are required. In the bootstrap procedure the original data set of data is considered as a parent population from which repeated samples are taken. The same number of nucleotides as the actual number used for constructing the tree are randomly sampled with replacement from the original data set, and a NJ tree is produced from this set of resampled nucleotide data. The topology of the tree is then compared with the original NJ tree and any interior branch of the NJ tree that

gives the same partition of sequences as that of the bootstrap tree is given a value of 1, whereas other interior branches are given 0. The process is repeated several hundred times and the percentage of times each interior branch of the NJ tree receives a value of 1 is computed. This bootstrap confidence limit is displayed on each branch and values greater than 70% are considered to be significant. MEGA was used to create distance matrices from the DNA sequences and Neighbour-Joining phylogenetic trees from these matrices.

With most studies of HIV involving PCR and cloning procedures there is the possibility of laboratory contamination leading to erroneous results and interpretations. To address this possibility, all newly generated sequences were initially included on a phylogenetic tree along with those of standard reference strains and local isolates. In all cases the branches containing sequences from a single individual were monophyletic, indicating that the contamination of samples with exogenous DNA had not occurred. This procedure also demonstrated that the patients used in this study were not superinfected with several viral strains of different origins, a situation that also confuses evolutionary analysis of HIV DNA sequences.

HIV-1 envelope sequences were assessed for inter-tissue recombination events using a new group-scanning method developed by Professor Peter Simmonds and described in detail in chapter 4. Nucleotide sequence alignments were edited in Simmonic to reduce alignment gaps and sequences labelled in groups according to tissue origin. Sequences were gap-stripped then bootstrap resampled J-C distance matrices and N-J phylogenetic trees for each window position were generated using

components within the Phylogeny Inference Package (PHYLIP: Felsenstein, 1989). A number of sequences were also assessed by the bootscanning method (Salminen et al., 1995) in the Simplot program (version 2.5). This method uses a sliding window moving in steps across the alignment. A tree is constructed for every window and bootstrap values calculated for the clusters. Values corresponding to the percentage of bootstrap resampled trees that a test sequence clusters with each user-defined group of sequences are plotted for each window position. This method was developed for the identification of inter-subtype HIV recombinant sequences, where the suspected recombinant sequence is compared with a representative sequence, consensus sequence, or set of sequences from each of the HIV-1 group M subtypes. In the region of the crossover point the bootstrap value first drops for the cluster that contains the recombinant. At some point after this, the sequence will move from its cluster into a new one, and the bootstrap value for this cluster should then increase as the sliding window moves past the crossover point. For *env* sequence analysis (Chapter 4) groups of sequences were assigned according to tissue origin (right occipital, left temporal, or lymph node), and test sequences sequentially removed from their group, bootscanned, then replaced. Bootscans were performed using as close parameters as possible to group-scan analysis, with a 500 bp window moving in 100 bp steps and trees generated with 50 bootstrap resamplings. As J-C distances are unavailable in Simplot Kimura 2 parameter distances were used to generate the distance matrix, with transversions scored at double the value of transitions.

2.17 Eukaryotic cell culture

2.17.1 Growth medium and Cell Lines

Cell lines used in this study are listed (Table 2.5). The 293T human embryonic kidney cell line was used for DNA transfection and NP2 (Soda et al., 1999) human glioma cell lines expressing different HIV receptors were used for infection with pseudotype virus. Cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) (Invitrogen) supplemented with 10% foetal calf serum (FCS, Sigma), 100 U/ml penicillin (Merck), 100 µg/ml streptomycin (Sigma), and 2 mM L-glutamine. Geneticin (G418, Life Technologies) and/or Puromycin (Life Technologies) was added to NP2 cells where appropriate. All cells were grown at 37°C in a humidified atmosphere containing 5% CO₂.

Cell Type	CD4/coreceptor expressed	Selective medium
293 T	-	-
NP2-CD4	CD4	1 mg/ml Geneticin
NP2-CD4/CCR5	CD4, CCR5	1 mg/ml Geneticin, 1 µg/ml Puromycin
NP2-CD4/CXCR4	CD4, CXCR4	1 mg/ml Geneticin, 1 µg/ml Puromycin

Table 2.5. Mammalian cell lines used for transfection experiments, HIV-1 receptors expressed and selective media required for maintenance of expression.

2.17.2 Thawing Cells

Frozen aliquots of cells were removed from liquid nitrogen storage and thawed rapidly at 37°C. 10 ml of cold media was then added dropwise to the cells in a 20 ml universal container. The universal was centrifuged at 500g for 5 minutes to pellet cells. The media was removed and the cells resuspended in 5 ml fresh media before

transfer into a 25 cm³ tissue culture flask. Cells were then checked daily and split/fed as required.

2.17.3 Passaging Cells

Cells were kept at less than 80% confluency at all times, split every 3-5 days at a ratio of 1:5-1:8, and discarded after 15 passages. Cells were grown in a 25 cm³ tissue culture flask and split by first removing the media and washing the monolayer with Phosphate Buffered Saline (PBS: 7.02 mM Na₂HPO₄, 1.47 mM KH₂PO₄, 0.137 M NaCl, 2.68 mM KCl) to remove FCS. Two ml of 0.02% versene (Sigma) was added and passed over the cells for 1 minute then removed and the process repeated with 2 ml of 0.05% trypsin (Invitrogen). The flask was incubated at 37°C until cells had detached from the plastic and 5 ml of fresh media was added to inactivate the trypsin. Cells were then diluted in media and added to a fresh tissue flask.

2.17.4 Freezing cells

Cells were pelleted by centrifugation at 500g for 5 min and resuspended in FCS containing 10% dimethyl sulphoxide (Merck) at a concentration of 10⁶-10⁷ cells/ml. 1 ml aliquots were dispensed into cryotubes (Nunc), which were placed in a freezing box and placed at -70°C (cooling cells at approximately 1°C per minute) overnight before transferring to liquid nitrogen for long-term storage.

2.18 Electroporation of 293 T cells

293 T cells were treated with trypsin and versene then suspended in media as in section 2.17.3. If at low concentration, cells were pelleted by centrifugation at 500g for 5 minutes and resuspended to a concentration of 10⁶ cells/ml. DNA for

transfection was mixed with 800 μ l of these cells and the mixture transferred to a 0.4 cm electrode gap Gene Pulser Cuvette (Biorad). The cuvette was placed in an electroporator (Biorad, Gene Pulser) and the cell/DNA mix subjected to a single pulse at 250V and 960 μ FD capacitance. Electroporated cells were then removed and mixed with 5 ml of media in a universal container. For EGFP and luciferase assays cells were then transferred to a single well in a six well tissue culture plate (Nunc), for *nef* assay 1 ml of diluted cells was added to a well in a 24 well tissue culture plate. Plates were incubated for 2 days before analysis.

2.19 Fluorescent Microscopy

293 T cells electroporated with EGFP gene were observed under an inverted fluorescent microscope 48 hours after transfection. EGFP within cells was observed by excitation at 488 nm resulting in emission of light at 507 nm. Pictures were taken using a Hamamatsu digital camera and then edited using the Openlab computer package.

2.20 Luciferase Assay

Expression of luciferase in 293 T cells was assessed 48 hours after electroporation of 293 T cells with *luc*⁺ gene. Luciferase activity was measured using components from the Dual-Luciferase Reporter Assay System (Promega). Cells were disassociated from the plastic of the six well tissue culture plate and resuspended in 1 ml media as in section 2.17.3. Cells were then transferred to a sterile eppendorf tube and pelleted by centrifugation at 6000g for 1 minute. The media was removed and the cells resuspended in 100 μ l of passive lysis buffer. The tube was then incubated for 15

minutes at room temperature to ensure complete lysis of cells. After incubation 10 μ l of the lysate was added to a 96 well plate, which was then placed in a Luminoskan RT luminometer. 25 μ l of Luciferase assay reagent II (containing the firefly luciferase substrate luciferin) was then mixed with the lysate and light production measured. The readings were blanked individually for each lysate prior to substrate addition in order to prevent strongly positive samples from creating artificially high readings in adjacent wells.

2.21 Nef Immunostaining

Expression of the electroporated *nef* linear expression construct in 293 T cells was detected by immunocytochemistry using components from the Streptavidin ABC Complex/HRP kit (Dako). A Nef specific monoclonal antibody (EVA3067.2 kindly provided by Dr K Krohn via the NIBSC Centralised Facility for AIDS Reagents) was added to monolayers, followed by a biotinylated rat-anti-mouse secondary antibody. The signal was amplified with a horse radish peroxidase labelled Streptavidin-Biotin Complex (ABC) and visualised by addition of the chromogen diaminobenzidine (DAB).

293 T cells were fixed 2 days after electroporation with water or *nef* LEC. Growth media was removed and cells washed with 500 μ l PBS. After removal of PBS cells were fixed by addition of 500 μ l 2% paraformaldehyde and incubated for 4 hours at room temperature. Paraformaldehyde was removed and cells washed once with 500 μ l 3 \times PBS and three times with 500 μ l 1 \times PBS. Cells were then blocked by addition of 200 μ l PBS containing 3% bovine serum albumin (BSA) and incubated at room temperature for 45 minutes. 200 μ l of primary antibody diluted 1/40 in PBS

containing 3% BSA was added to each well and incubated for 1 hour at room temperature. Cells were then washed twice with 500 μ l of PBS then 200 μ l of biotinylated secondary antibody diluted 1/200 with PBS + 3% BSA was added and the plate incubated for a further 45 minutes at room temperature. After incubation with secondary antibody cells were washed twice with 500 μ l PBS then 200 μ l of streptavidin-biotin complex (ABC) was added to each well. Cells were washed twice with PBS and 200 μ l of DAB (Vector laboratories) was added to each well. The plate was then incubated at room temperature until brown foci developed (approximately 5 minutes). Cells were washed twice with 500 μ l of distilled water and 200 μ l of distilled water was added to each well. Photographs were taken under a light microscope at 40x magnification and edited using the Corel Photo-Paint 8 computer package.

2.22 Pseudotype Virus Production and Analysis

Several control clones were required for the production of pseudotype virus (Fig 2.2). The viral proteins needed were provided by the HIV backbone clone pNL4.3 Δ envGFP (Fig. 2.2.C), a construct based upon the tissue culture adapted, syncytium inducing, CXCR4 using strain NL4.3.

This construct has a defective *env* gene produced by filling in the *Nde*I site indicated, therefore causing a frameshift (He et al, 1997). The backbone clone also has green fluorescent protein inserted into *nef*. The positive control *env* expressing clones used in this study were YU2 *env* (derived from a CCR5 using virus), and HXB2 *env* (derived from a CXCR4 using virus) (Fig. 2.2.B). These two control *env* constructs had previously been introduced into the pSVIII vector (Choe et al, 1996).

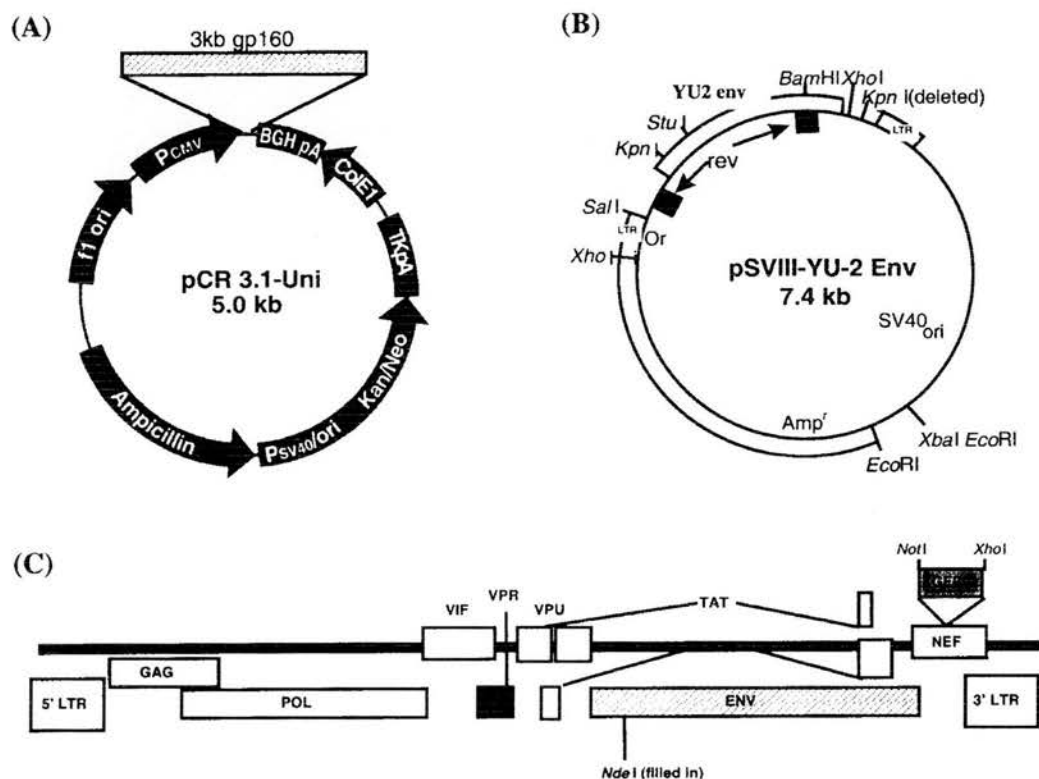


Figure 2.2. Vectors used for pseudotype virus production. **A)** *env* gene cloned into pCR3.1-unidirectional vector. **B)** The CCR5 positive control *env* clone YU2 in the vector pSVIII. **C)** pNL4.3 Δenv GFP proviral backbone clone (in the pSVIII vector). When co-transfected with an Env expressing clone this *env* defective construct provides the additional HIV proteins required for virion formation. In addition to a frameshift in *env* it also contains GFP in place of *nef*, which is expressed upon infection of a new cell with progeny pseudotype virus.

All control clones used in this study were kindly provided by Dr Paul Clapham. Patient derived envelopes were either cloned into pCR3.1 (Fig. 2.2.A), or generated by ligation into this vector with subsequent amplification of linear expression constructs (Section 2.10).

For pseudotype virus production an *env* expressing genetic element (either a clone or a linear expression construct) was cotransfected along with an *env* deficient proviral backbone into 293T cells by calcium phosphate transfection. The cells were incubated for 48 hours then the supernatant harvested and filtered (0.45 μ m) to remove cellular debris. Serial 10-fold dilutions of harvested virus were added to NP2 indicator cells expressing either CD4 alone, or in conjunction with CCR5 or CXCR4. After incubation for 48 hours infection of these cells was detected directly by observing GFP expression with fluorescent microscopy or indirectly by p24 immunocytochemistry.

2.23 Calcium Phosphate Mediated Transfection:

Prior to transfection, DNA was ethanol precipitated. Ten μ g of pNL4.3 Δ *env*GFP was mixed with 4 μ g of envelope gene (plasmid or LEC), 0.1 volume of sodium acetate (pH 4.5) and 2.5 volumes of cold (-20°C) 100% ethanol in a 1.5 ml eppendorf tube. This mixture was centrifuged at 12000g for 30 min at 4°C, the supernatant removed, and the DNA pellet washed by addition of 125 μ l 70% ethanol. The tube was centrifuged again at 12000g for 5 minutes at 4 °C, the ethanol removed, and DNA air dried for 5 minutes. DNA was then resuspended in 263 μ l sterile water.

Calcium phosphate transfections were performed using the ProFection Mammalian Transfection System - Calcium Phosphate (Promega). 293 T Cells were plated out the day before transfection onto a 6 well tissue culture plate at a density of 10^6 cells/well. This seeding density provided a monolayer of cells at approximately 60% confluency on the day of transfection. Three hours prior to transfection the media was removed from cells and replaced with 5 ml fresh growth medium. For each transfection the DNA and $2 \times$ HEPES-Buffered saline (HBS) solutions were prepared in separate sterile tubes. In the first tube the DNA was mixed well with 37 μ l of 2M CaCl_2 . Working in a tissue culture hood the second tube containing 300 μ l of $2 \times$ HBS was vortexed gently at a speed allowing addition of the prepared DNA solution. Whilst continuing to vortex the CaCl_2 /DNA solution was slowly added dropwise. The solution was then incubated at room temperature for 30 minutes. After incubation the transfection solutions were vortexed and added dropwise to 293 T cells. The cells were then incubated for 24 hours, and the media containing CaPO_4 removed and replaced with 5 ml fresh DME media. Cells were incubated for a further 24 hours before harvesting of pseudotype virus. Media was removed from cells and filtered through a 0.45 μ M filter (Nunc) to remove any floating cells and cellular debris. Several 1 ml aliquots were then transferred to cryotubes and stored at -70°C before use. This pseudotype virus was then used to infect NP2 cells to detect production of functional virus and analyse coreceptor usage.

2.24 Viral Infectivity Assays

NP2 cells were seeded the day before infections at a density of 1.5×10^4 cells/well in a 48 well tissue culture plate (Costar). Frozen aliquots of virus were thawed at room

temperature and DME media used to set up 10-fold serial dilutions (from undiluted to 10^{-5}). The media was removed from NP2 cells and 100 μ l of each dilution of virus was added to a separate well in the tissue culture plate. The plate was incubated overnight (16 hours) then 400 μ l of DME media was added to each well (without removal of virus). Plates were incubated for a further 48 hours and analysed by fluorescent microscopy (as in section 2.19) or immunostaining using anti-p24 monoclonal antibodies.

2.25 p24 immunocytochemistry

Cells infected with pseudotype virus were stained with a 1:1 ratio of mouse anti HIV-1 p24 monoclonal antibodies EVA365 and EVA366 kindly provided by Dr B Wahren via the NIBSC Centralised Facility for AIDS Reagents. Primary antibodies were stained with goat anti-mouse β -galactosidase conjugated monoclonal antibody (Southern Biotechnology). p24 positive cells were then visualised by using the β -galactosidase reporter gene staining kit (Sigma).

Three days after addition of virus, media was removed from wells and the cells fixed by addition of cold (-20°C) methanol:acetone (1:1). After 2 minutes methanol/acetone was removed and 500 μ l of PBS containing 1% FCS was added to each well. PBS was removed and 100 μ l of anti-p24 antibodies diluted 1/200 in PBS containing 1% FCS was added. Cells were incubated for 1 hour at room temperature with anti-p24 antibody. The cells were then washed twice by aspirating off liquid from cells and adding 500 μ l of PBS containing 1% FCS. The second wash was removed and 100 μ l of goat anti-mouse β -galactosidase conjugated antibody diluted

1/400 in PBS containing 1% FCS was added to each well. The plate was incubated again for 1 hour at room temperature then washed 3 times in PBS only. PBS was removed and 200 μ l of substrate (0.5 mg/ml 5-Bromo-4-chloro-3-indolyl- β -D-galactopyranosidase (X-gal), in PBS containing 1 mM MgCl₂, 3 mM K₃Fe (CN)₆, 3 mM K₄Fe (CN)₆) was added to cells. The plate was then incubated at 37°C until blue foci developed (approximately 1 hour). Plates were then washed once with PBS and 200 μ l PBS was added to each well before storage at 4°C. Foci of infection were counted by eye under light microscopy at 20 \times magnification. Focal forming units present in the initial harvested media was calculated by taking the average number of foci of infection present in the last two positive wells and multiplying by the dilution factor.

Chapter 3: Results (i)

3.1 Introduction

Central nervous system abnormalities are common among HIV-1 infected individuals (Snider et al., 1983). Approximately 80% of AIDS patients experience some form of neurological or psychiatric symptoms and large post mortem studies have found neuropathological abnormalities in 79-94% of brains from individuals who died of AIDS (Burns et al., 1991; Bell, 1998). Clinical presentations include motor and cognitive deficits along with behavioural changes and other symptoms that eventually progress to HIV associated dementia (HAD) in approximately 10%-30% of untreated infected individuals (Price, 1996; Bell et al., 1998; Gray et al, 1996). This manifestation of disease is therefore a major cause of morbidity in AIDS patients and may develop as the predominant clinical abnormality at advanced stages of HIV-1 infection (Navia and Price, 1987).

HIV encephalitis, often characterised by the presence of multinucleated giant cells in the brain, is observed in approximately 10%-50% of AIDS autopsies (Shaw et al., 1985; Bell, 1998). The major cellular reservoir for HIV-1 at this site is monocyte derived macrophages and microglial cells (Wiley et al, 1986; Bagasra et al, 1996), although a number of other cell types may also be infected at lower levels (Bagasra et al, 1996; Ranki et al., 1995; Wiley et al, 1986; Bell et al., 1993; Gyorkey et al, 1987). HIV-1 induced damage to neuronal cell types can occur by at least two mechanisms; shedding of toxic viral proteins by infected cells (Kaul and Lipton, 1999; Lipton, 1993; Nath et al., 1996; Sabatier et al, 1991), and secretion of inflammatory cytokines and other molecules by infected or activated macrophages and microglial cells (Lipton and Gendelman, 1995; Giulian et al., 1990; Yeh et al.,

2000). In addition to the pathological effect that HIV-1 can have within the brain, the poor penetration of several antiretroviral agents into the CNS (Enting et al., 1998) has led to concerns that it may serve as a “sanctuary site”, allowing continued virus replication despite effective suppression elsewhere in the body. As the lifespan of HIV-1 infected individuals is extended by the use of antiretroviral drugs, this continued virus replication in the CNS could be sufficient to instigate a neurological decline culminating in dementia. Moreover the presence of sub-optimal drug concentrations in the CNS may provide an ideal environment for the evolution of drug resistant HIV-1 variants, which could then re-seed the peripheral blood and lymphoid system.

A central question in determining the importance of the CNS as an anatomical reservoir for HIV-1 concerns the time that this site is stably infected. One possibility is that at some point before the onset of AIDS, HIV-1 establishes an autonomous, self-sustaining infection of the CNS that is not dependent upon continued replenishing by newly infiltrating infected cells. HIV-1 in the brain may therefore be outwith the reach of current antiviral therapies, and concerns over the role of the CNS as a sanctuary site could be well founded. A second possibility is that autonomous replication of HIV-1 within the brain is only established at, or around, the onset of AIDS. If this is true then effective suppression of virus in the peripheral blood and lymphoid organs during the asymptomatic period of infection should be sufficient to prevent complications associated with HIV-1 in the brain. These early and late entry models each have supporting evidence, and determining the relative contribution of each is important for enhancing our understanding of HIV-1 within the CNS.

As mentioned, an early entry model suggests that the brain is stably infected before the onset of AIDS, possibly soon after transmission. Low levels of HIV replication or viral latency would then follow, before eventual reactivation of HIV in this site, perhaps facilitated by the general immunosuppression associated with the onset of AIDS. In support of this model is the detection of HIV proviral DNA in the brains of seropositive but asymptomatic individuals who died of AIDS-unrelated causes (Gray et al, 1996; An et al., 1996), and the appearance of HIV in the brain within 2 weeks of accidental iatrogenic infection (Davis et al, 1992; Palmer et al, 1994). These observations have been confirmed in the closely related simian immunodeficiency virus (SIV)/macaque model, where SIV can be detected in the brains of infected animals within 2 weeks of peripheral inoculation (Smith et al., 1995). A late entry model suggests that large-scale autonomous viral replication in the brain is the result of single or multiple seeding of the CNS from the periphery during AIDS, when there is a high viral load and the integrity of the blood brain barrier may be compromised (Dallasta et al., 1999). This hypothesis is supported by the fact that HIV is absent before the onset of AIDS, and severe neurological impairment invariably occurs late in infection (Price et al., 1988; Bell et al, 1993; Gray et al, 1992). HIV-1 infected cells in the brain are also frequently perivascular in location, suggestive of recent immune infiltration (Vazeux, 1991).

A second issue concerning HIV-1 in the brain is the frequently observed genetic segregation between viral variants present in the brain and lymphoid system (Korber et al, 1994a; Ball et al, 1994; Di Stefano et al, 1996; Epstein et al, 1991b; Wong et al, 1997; Vantwout et al, 1998; Hughes et al, 1997b), usually based upon partial *env* sequences. Factors that could lead to the development of a separate virus

subpopulation in the brain include the original seeding of this site being followed by a long period of physical isolation between brain and lymphoid variants. This would amount to a founder event, allowing the two populations to diverge by genetic drift. Another possibility is that virus in the brain has a lower turnover rate than in the lymphoid system, due for example to a relative lack of potential host cells or longer periods of latency. Brain-derived virus may then consist of archival sequences, the remnants of previously abundant viral variants present in the lymphoid compartment. The relatively immunoprivileged nature of the CNS may also lead to comparatively weak immune-mediated selective sweeps against particular viral variants, possibly allowing more genetic drift to occur. Alternatively, the genetic segregation may be a result of adaptation to the particular microenvironment and available host cells within the brain. If this is the case then lymphoid variants may enter the brain relatively frequently, but be poorly equipped for prolonged survival in the unique environment of the CNS.

A final and more complex hypothesis is that viral variants present in different anatomical sites of the body, such as the colon (Wang et al, 2001) or bone marrow (Gartner and Liu, 2002; Gartner, 2000; Liu et al., 2000), have diverged from those in the lymphoid system because of circumstances particular to those tissue sites, and subsequently entered the brain. One possible reason for this divergence may be the development of HIV-1 variants specifically adapted for infection of tissue macrophages as opposed to CD4⁺ T lymphocytes (Wang et al, 2001). Virus that has adapted to infect these cells may then also be pre-adapted for infection of macrophages and microglia within the brain. Trafficking of virus or virally infected cells from these sites into the brain could then seed an infection with the genetically

distinct virus subpopulation that evolved elsewhere in the body. If this is the case then the putative anatomical site that the genetic divisions emerged is shifted, but the question of which underlying factors are responsible for the outgrowth of distinct viral subpopulation within the brain remains.

While direct evidence for recombination between HIV-1 variants that have diverged during a single infection is currently limited (Jung et al, 2002), the large number of reports describing recombination between more divergent HIV-1 strains suggests that the former may be a common event (Carr et al, 1998; Bobkov et al, 1998; Liitsola et al, 1998; Liu et al, 2002; Diaz et al, 1995). Therefore one potential means of assessing the underlying factors leading to apparently tissue-specific HIV-1 populations is to look for evidence of recombination between viral variants in different compartments. If recombination has occurred then physical isolation or differences in replication rates are unlikely to have given rise to the separate populations. Moreover, the observation of recombination in this case would also provide more general insights into the adaptive and evolutionary mechanisms associated with HIV-1 compartmentalisation. Recombination can be detected by the presence of phylogenetic relationships between sequences that vary depending upon the particular region of the viral genome analysed. This may include the observation of clearly discordant phylogenetic relationships based upon different subgenomic regions, or a marked reduction in sequence diversity in one subgenomic region that is not apparent elsewhere in the viral genome (Brown and Cleland, 1996).

Therefore in this study the genetic relationships manifest in different HIV-1 subgenomic regions were assessed, both within brain and between brain and

lymphoid sequences. The two subgenomic regions chosen for this analysis were p17^{gag} and V3. p17^{gag} sequences have previously proved useful in ascertaining the evolutionary relationships of HIV-1 variants both between (Holmes et al., 1993; Kasper et al., 1995) and within (Hughes et al, 1997b; Wang et al, 2001) HIV-1 infected individuals. The investigation of haemophiliacs infected from a common batch of clotting factor concentrate has allowed the rate of sequence change in p17^{gag} to be calculated (Kasper et al, 1995). This in turn may be used to calculate approximate times of divergence for any pair of sequences. Thus if p17^{gag} sequences were found to segregate between brain and lymphoid variants, then an approximate divergence time of the two viral subpopulations could be calculated, providing information on timing of infection of the brain. The V3 region encoding the V3 loop of gp120 has been implicated as the major determinant of several viral properties such as cell tropism (Hwang et al, 1991), coreceptor preference (Speck et al, 1997), ability to produce syncytia (Fouchier et al, 1992; Fouchier et al., 1995b), and as the principal neutralising epitope of host humoral responses (Wolfs et al., 1990; Palker et al., 1988; Rusche et al., 1988). HIV-1 V3 sequences have also been found to segregate with some frequency between brain and lymphoid compartments in HIV-1 infected individuals (Korber et al, 1994a; Chen et al, 2000; Epstein et al, 1991b).

3.2 Aims

- 1) Amplify and sequence the V3 and p17^{gag} regions of the HIV-1 genome using template DNA from lymphoid and brain tissue of infected individuals with evidence of HIV encephalitis.

2) Compare the sequence diversity and phylogenetic relationships of virus present in brain and lymphoid tissue in the two subgenomic regions and where possible calculate approximate divergence times based upon p17^{gag} distances.

3) Search for evidence of discordant phylogenetic relationships (indicative of recombination) in the different genome regions between virus derived from separate tissue sections.

3.3 Clinical details of study subjects

Samples of LN or spleen and several anatomically distinct regions of brain from four individuals with HIV giant cell encephalitis were stored at -40°C at autopsy. Risk group, CD4 counts, and brain pathology of study subjects are summarised (Table 3.1). Subject NA128 had received zidovudine monotherapy for 17 months up to approximately 1 year before death; ddC was used for 1 month, finishing 3 months before death.

The other study subjects had received minimal antiviral treatment: for NA234, a single course of zidovudine for a duration of 1 month at 1 year before death; for NA021, zidovudine intermittently over 1 year at 5 years before death and zidovudine-ddC for 1 month at 1 year before death; for NA173, zidovudine for 4 months at 2 years before death and then for 1 month at 1 year before death. None of the study subjects showed evidence for genotypic resistance to zidovudine or other antiviral agents (P Strappe pers comm). The brains were examined pathologically as previously described (Bell et al, 1993). Assessment of pathology findings was undertaken blind to the PCR analysis and validated by three independent observers.

Study subject	Age (yr)	Risk group ^a	CD4/ μ l		HIV-related pathology ^b						Other pathology
			1 yr before death	3 mo before death	PV	WM	GM	BG	BS	Cere	
NA173	34	MH	1	0	++	+/-	+/-	+/-	+/-	-	Toxoplasma
NA021	49	IVDU	300	87	+	++	+	+	+/-	+/-	
NA128	48	MH	12	4	+/-	+++	++	++	++	-	Toxoplasma
NA234	28	IVDU	90	30	+	++	++	++	+	+	Lymphoma

Table 3.1. Clinical background and pathology appearance of study subjects. ^aMH, male homosexual; IVDU, intravenous drug user. ^bScored as – (no pathology), +/- (sparse pathology), + (slight pathology), ++ (medium pathology), or +++ (severe pathology in different brain compartments). PV, perivascular; WM, white matter; GM, grey matter; BG, basal ganglia; BS, brain stem; Cere, cerebellum.

Comparison	Study subject	Distance ^a					
		p17 ^{gag}			V3		
		Mean	Error	Rank ^b	Mean	Error	Rank
Within brain	NA021	0.034	0.009	3	0.015	0.008	4
	NA173	0.035	0.010	2	0.017	0.008	3
	NA128	0.012	0.006	4	0.035	0.011	1
	NA234	0.042	0.011	1	0.028	0.010	2
Brain-lymphoid tissue	NA021	0.040	0.009	3	0.043	0.014	3
	NA173	0.040	0.011	2	0.031	0.011	4
	NA128	0.023	0.008	4	0.080	0.018	1
	NA234	0.057	0.013	1	0.072	0.018	2

Table 3.2. p17^{gag} and V3 sequence divergence. ^aMean J-C distance between variants within brain or between brain and lymph node or spleen (excluding choroid plexus). ^bRank of variability (1 = most variable).

3.4 Methods

A full description of methods is provided (Chapter 2). In brief, DNA was extracted from each sample by the phenol/chloroform method and quantified spectrophotometrically. HIV proviral DNA was semiquantitated by nested PCR amplification using the p17^{gag} primers on serial 10-fold dilutions of DNA, with the last positive dilution used to indicate the minimum proviral load in the sample. Samples were used for sequence comparison only if proviral frequencies were >100 copies/10⁶ cells. Low proviral levels were detected in two atrophied LN samples from NA128, and lymphoid sequences were therefore obtained from the spleen. p17^{gag} and V3 regions were then amplified by nested PCR using 1 µg aliquots of DNA as template and directly sequenced manually. Alison Morris, a co-worker in the laboratory, provided all sequences from NA234 by cloning and sequencing of single or multiple clones from each region as described in (Morris et al., 1999). Phylogenetic analysis was performed with the MEGA program (Kumar et al., 1993). Each set of sequences from the four study subjects was monophyletic in both genomic regions and distinct from those of the published sequences of subtype B (GenBank accession numbers in parentheses): HIVSF2 (K02007), HIVRF (M17451), HIVOYI (M26727), HIVLAI (K02013), HIVJRFL (M74978), HIVYU2 (M93258), HIVCAM1 (D10112), HIVNY5CG (M38431), HIVHAN (U43131), HIVWMJ22 (M12507), and HIVSFAAA (M65024). This comparison provided no evidence for coinfection with more than one epidemiologically unrelated HIV strains or for intersample or exogenous laboratory contamination. Previous sampling problems associated with using a single small tissue section to reflect the total virus population in a large solid organ (Hughes et al, 1997b) were partially overcome by

using several anatomically distinct regions of brain for sequence analysis. Various samples were also taken from both left and right areas of brain to allow comparison of virus present in areas that have equivalent cellular components but are physically separate.

3.5 Results

3.5.1 V3 and p17^{gag} tree topology

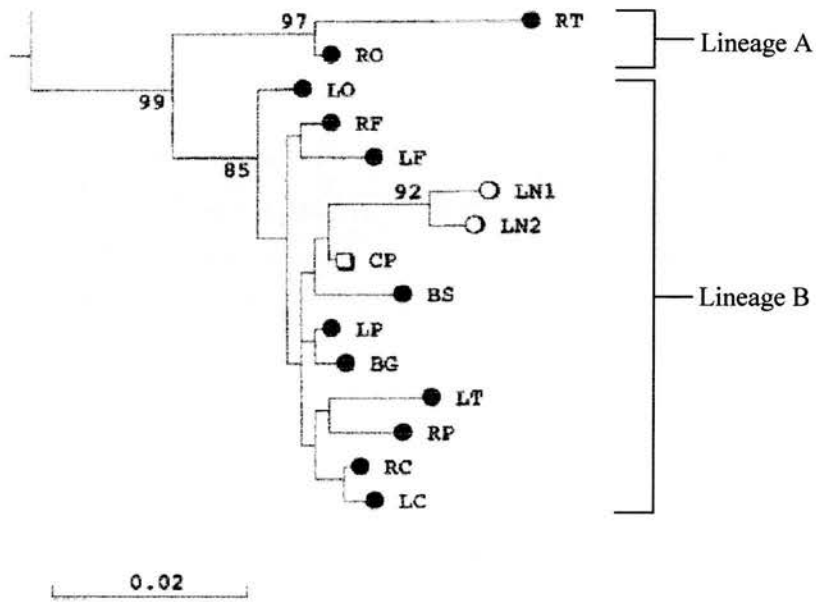
NA021

The p17^{gag} tree topology for subject NA021 displayed three bootstrap supported lineages (Fig. 3.1.A). Lineage A contained only the sequences from the right temporal and right occipital lobes, and lineage B contained sequences from the remainder of brain samples and the two lymph nodes.

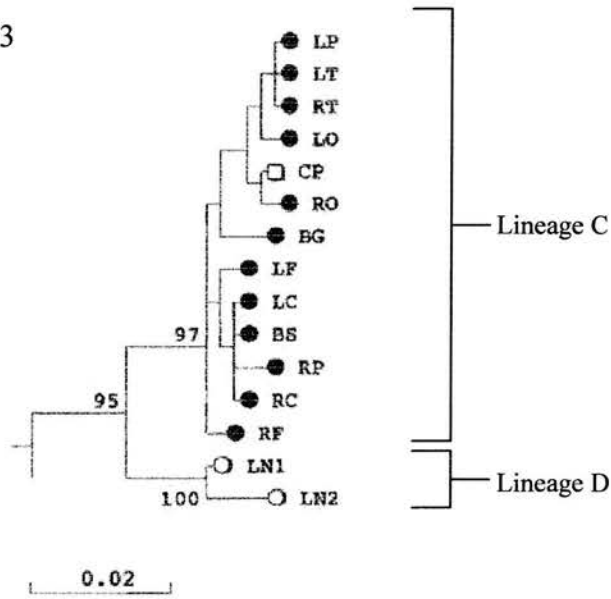
The third, less pronounced lineage demonstrated that the two lymph node sequences were more closely related to each other than to the remaining sequences in lineage B. The mean synonymous J-C pairwise distance between lineages A and B was 0.0585 (mean standard error +/- 0.0270). On the basis of the previously established mean rate of sequence change in this region of *gag* (0.6 to 0.7 % per site per year: Hughes et al, 1997b; Kasper et al, 1995), these distances suggested a time of divergence between lineages A and B of 4.43 years +/- 2 years, compared to the onset of AIDS in this subject of 2 years before death and a CD4 count of 300 cells/ μ l one year before death. This implied that at least some of the virus (within the RO and RT regions) present in the brain of this study subject during late stage HIV infection originally diverged from the lymphoid virus population before the individual progressed to AIDS.

Figure 3.1. Phylogenetic analysis of p17^{gag} and V3 sequences from subject NA021. A) p17^{gag} phylogenetic tree, B) V3 phylogenetic tree. Sequences available from GenBank listed in Methods were used to root each clade. Neighbour-joining trees were plotted using the indicated scale of J-C distances. All bootstrap values $\geq 75\%$, generated with 100 resamplings, are indicated on branches. Labels indicate sequences from LN: lymph node, RT: right temporal, LT: left temporal, RO: right occipital, LO: left occipital, RF: right frontal, LF: left frontal, CP: choroid plexus, BS: brain stem, RP: right parietal, LP: left parietal, BG: basal ganglia, RC, right cerebellum, LC: left cerebellum. Symbols indicate sequences derived from brain (●), from lymph node (○), or from choroid plexus (□).

A) p17^{gag}



B) V3



V3 sequences derived from the same DNA samples produced a tree containing two bootstrap supported lineages (Fig. 3.1B). In marked contrast to the p17^{gag} tree, lineage C contained all brain-derived sequences, and lineage D contained only the sequences derived from the two lymph nodes. Virus from the right occipital and right temporal lobes was therefore distinct from other brain sequences in p17^{gag} but closely related in V3, with the converse being true for lymph node-derived virus. One interpretation of this finding is that the virus found in the right temporal and right occipital lobes represents a pre-AIDS seeding of the brain, with the remainder of the brain regions being infected later by the predominant virus in the lymphoid system at that time. The observed similarity in V3 sequences among brain isolates may then subsequently have arisen by strong convergent evolution. However, both synonymous and non-synonymous mutations in V3 were shared by brain but not lymphoid sequences (Fig. 3.5.A, and data not shown). Therefore recombination between brain- and lymphoid-derived virus is the most likely explanation for the observed discordant phylogenies.

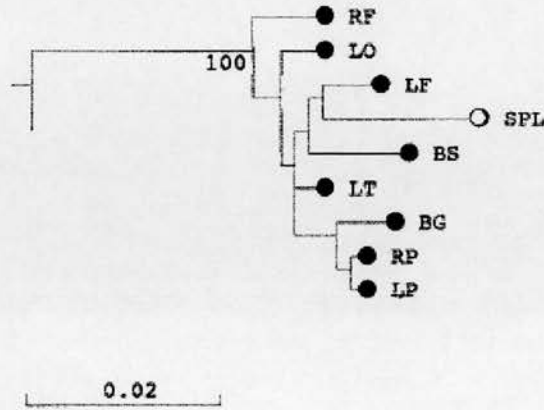
NA128

For subject NA128 the p17^{gag} relationships were monophyletic (Fig. 3.2.A). However the V3 phylogenetic tree exhibited a pronounced bifurcation (Fig. 3.2.B), with Lineage A containing sequences derived from left and right frontal lobes along with that from lymphoid tissue (spleen), and Lineage B containing the remainder of brain sequences.

NA173

p17^{gag} sequences from study subject NA173 produced a tree with three bootstrap supported lineages (Fig. 3.3.A).

A) p17^{gag}



B) V3

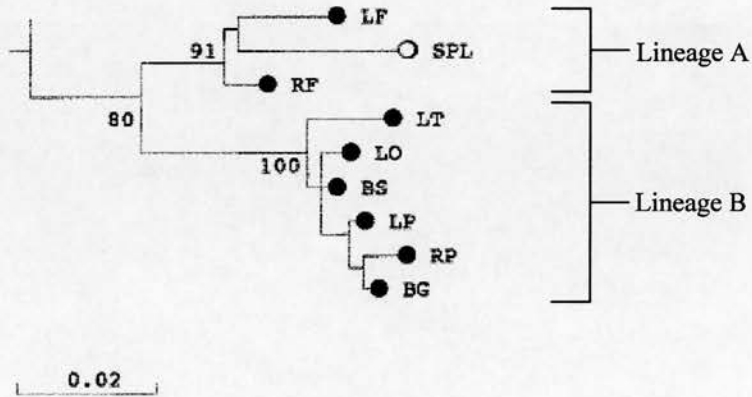
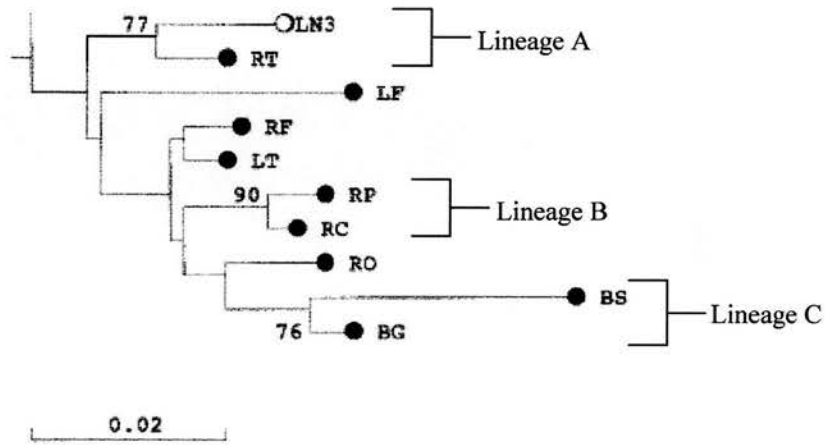


Figure 3.2. Phylogenetic analysis of p17^{gag} and V3 sequences from subject NA128. A) p17^{gag} phylogenetic tree. B) V3 phylogenetic tree. Outgroup, labels, symbols, and bootstrap method as for Fig 3.1. Lymphoid sequences were in this case obtained from spleen (SPL).

A) p17^{gag}



B) V3

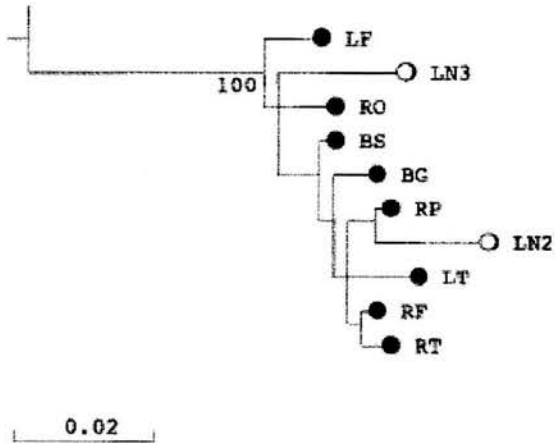


Figure 3.3. Phylogenetic analysis of p17^{gag} and V3 sequences from subject NA173. A) p17^{gag} phylogenetic tree. B) V3 phylogenetic tree. Outgroup, labels, symbols, and bootstrap method as for Fig 3.1.

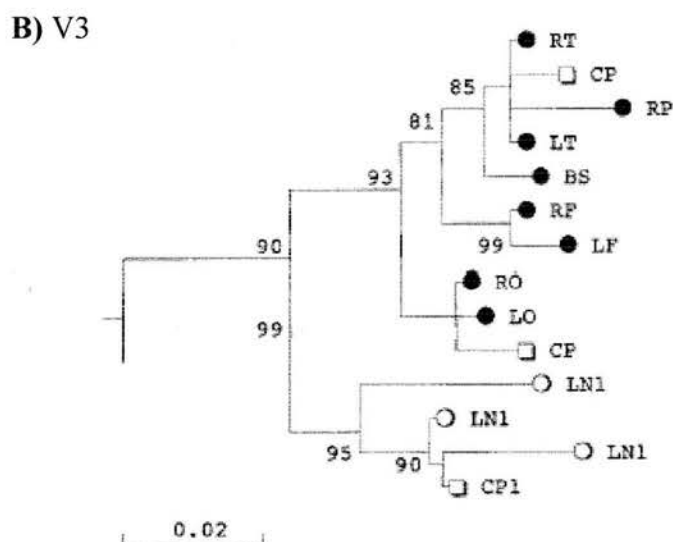
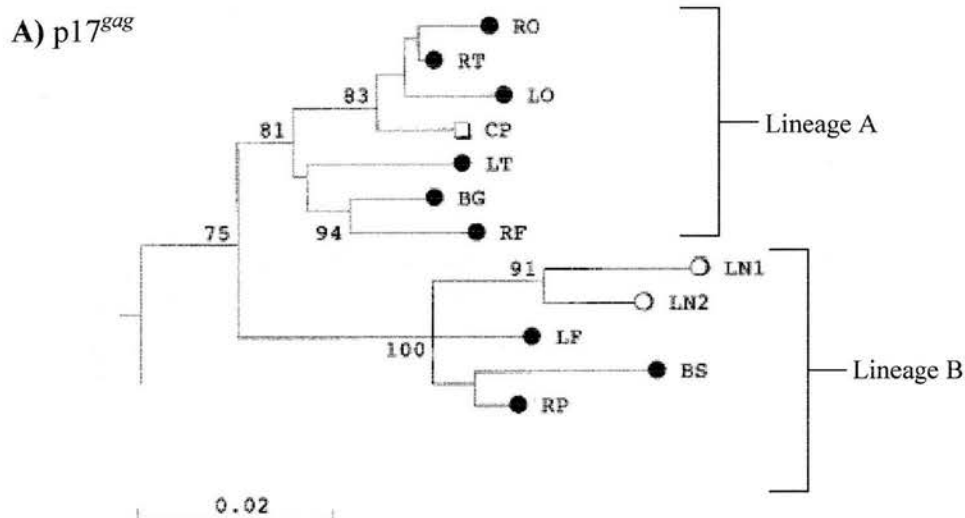


Figure 3.4. Phylogenetic analysis of p17^{gag} and V3 sequences from subject NA234. A) p17^{gag} phylogenetic tree. B) V3 phylogenetic tree. Outgroup, labels, symbols, and bootstrap method as for Fig 3.1. Sequences from NA234 were obtained from individual clones of amplified DNA rather than by direct sequencing; single representative clones from each sample or multiple clones for those containing sequences in more than one lineage (i.e., LN and CP) have therefore been included.

Lineage A contained the lymph node and right temporal sequences, lineage B contained sequences from the right parietal and right cerebellum regions, and lineage C contained sequences from the basal ganglia and brain stem. The remainder of sequences was interspersed between these supported lineages. This topology is suggestive of several independent infiltrations into the brain from the periphery. As for the other study subjects, V3 sequences from this individual displayed a different tree topology than that of p17^{gag}, being monophyletic with no bootstrap supported clades present (Fig. 3.3.B).

NA234

The phylogenetic tree constructed with p17^{gag} sequences from study subject NA234 had two major supported lineages (Fig. 3.4.A), and several supported divisions within these lineages. Sequences in lineage A were solely brain derived and virus in these anatomical sites had common ancestors distinct from those in lymphoid tissue, more recent times of divergence can be inferred. Variants within lineage A may therefore have originated from the spread of HIV within the brain rather than from multiple seedings from the peripheral circulation and thus differ in origin from brain derived variants in lineage B (left frontal, right parietal, and brain stem). The mean synonymous J-C pairwise distance between the two lineages was 0.1098, suggesting a divergence time of 8.32 years +/- 2.8 years, compared with the individual progressing to AIDS approximately 3 years before death. The V3 phylogenetic tree displayed similar topology to that of study subject NA021, with lymphoid sequences grouping separately from all brain derived sequences with the exception of some choroid plexus sequences, including those that grouped with LN sequences in the p17^{gag} region (Fig. 3.4.B).

3.5.2 Sequence divergence and pathology appearance

The severity of HIVE varied between the study subjects (Table 3.1), ranging from infrequent giant cells confined to perivascular regions (NA173) to widespread pathology affecting both white and grey matter (NA128). The extent and type of HIV-induced pathology correlated with the degree of V3 but not p17^{gag} sequence diversity between brain regions and in the extent to which V3 sequences from the brain differed from those in lymphoid tissue. The mean J-C distances for V3 and p17^{gag} sequences both within brain and between brain and lymph node derived virus are shown (Table 3.2). V3 sequences from NA173 (who showed minimal HIV-related pathology) were largely undifferentiated from those detected in lymph node, with only an alanine-arginine change segregating by tissue (Fig. 3.5). At the other extreme, the spleen-derived sequence from NA128 had a predicted syncytium-inducing phenotype and differed in multiple sites from the nonsyncytium-inducing variants in the brain. However, in the p17^{gag} region sequences from NA173 were the most variable, and those from NA128 were the least variable.

3.6 Discussion

3.6.1 Timing of viral entry into the CNS.

This study documents the complex genetic interrelationships of HIV-1 populations *in vivo*, findings that have several implications in understanding mechanisms of tissue adaptation and the timing of entry of HIV-1 into the CNS. The different sequence relationships between brain-derived and lymphoid variants from the four study subjects suggest that entry of HIV-1 into the CNS can occur at different times. Clear genetic segregation in p17^{gag} sequences from subjects NA021 and NA234 between

lymphoid and non-lymphoid containing lineages allowed the mean synonymous J-C pairwise distances and therefore approximate divergence times to be calculated. In both cases the time since divergence of the two lineages pre-dated the individuals progression to AIDS, implying that the virus predominating in the brain during late stage HIV-1 infection may have originally entered the CNS before the subject advanced to AIDS. Intrinsic to this type of calculation is the assumption that the virus sampled has not undergone significant periods of latency before reactivation, although if this had occurred then it would result in an underestimation of viral divergence times. The interpretation of a pre-AIDS entry of HIV-1 into the CNS of these two subjects as predicted by an early entry model of HIV-1 infection of the brain is therefore still valid. A more complex proposal that viral variants present in the brain originally diverged from those in the lymphoid compartment at a third anatomical site before trafficking into the CNS (Gartner and Liu, 2002; Gartner, 2000; Liu et al, 2000) cannot be corroborated or discounted on the basis of this data. Therefore the possibility that the time since divergence of variants in the brain and lymphoid compartment does not correspond to the time of HIV-1 invasion of the brain remains open.

The poor penetration of antiretroviral agents such as protease inhibitors and some nucleoside analogues into the CNS (Lewis et al., 1996; Yarchoan et al., 1988) make the presence of this replication competent viral reservoir in the brain problematic for several reasons. Firstly the life expectancy of HIV infected individuals HAART may be greatly extended. However it is not yet clear whether the persistence and possible replication of HIV in the CNS of these individuals will eventually lead to the neurological symptoms generally associated with progression to AIDS. The presence

of sub-optimal concentrations of antiviral drugs in the CNS also provides ideal conditions for the development of antiviral resistance mutants, perhaps leading to re-seeding of the lymphoid system with drug resistant virus from the brain and failure of antiviral therapy. Finally, one potential aim of therapeutic interventions against HIV is the complete eradication of virus from infected individuals, ultimately allowing cessation of therapy. The presence of anatomical reservoirs containing replication competent virus such as the CNS represents a significant barrier to these endeavours.

Evidence for the contribution of late-entering variants to HIVE is provided by sequences from NA173, where a distinctive pathology appearance comprising HIV-expressing infiltrating macrophages confined to the perivascular regions was evident (Table 3.1). The hypothesis of recent entry of HIV-infected cells into the brain parenchyma was supported by the observation of close sequence similarity in the V3 region of brain derived variants with those obtained from lymphoid cells (Table 3.2. and Fig. 3.3.B). This late entry picture contrasted strongly with the distribution of HIV infection in NA128, in which HIV was widely dispersed in white and grey matter, while V3 sequences were largely distinct between spleen and brain and heterogeneous within brain (Table 3.1, and Fig.3.2). However this correlation was not supported by sequence comparisons in the p17^{gag} region where sequence diversity was greatest in NA173 and least in NA128 (Table 3.2). The increased within-brain heterogeneity of V3 sequences in putative early-entry variants could possibly be the result of a less-vigorous humoral immune response within the CNS allowing increased genetic drift.

One possible explanation for these contrasting results regarding timing of entry into the brain in different individuals is that HIV-1 invasion of this organ is similar to that proposed for HIV in CSF (Price, 2000). Where it was suggested that during the asymptomatic period virus in CSF infiltrates from blood in a "transitory" infection and as the individual progresses to AIDS a more "autonomous" infection in the CSF is established. In some cases, such as in subjects NA021 and NA234, this transitory infection may be sufficient to establish a persistent infection in the brain before the onset of AIDS. In others, such as NA173, an autonomous infection is not established until the onset of AIDS.

3.6.2 Identification of *in vivo* recombination events in the evolution of regional viral populations.

When sampling gene sequences from an asexually reproducing population where recombination has not occurred, the evolutionary relationships exposed by analysis of one gene would usually be representative of those determined using other genes from the same organisms. If the population is sexually reproducing or exchanges genetic information by recombination, then different genes from the same organism can have distinct evolutionary histories. In this study the different degrees of variability and the discordant phylogenetic relationship between p17^{gag} and V3 regions in each study subject indicated a lack of genetic linkage between these two subgenomic regions and supports the hypothesis of frequent recombination *in vivo*. The differing relationships between samples in the p17^{gag} and V3 regions also imply that the frequently observed tissue segregation in regions of the *env* gene is not solely the result of compartmentalisation or differences in viral turnover rates in the CNS compared with that in the lymphoid tissue. If this were the case then sequence

relationships in other regions of the viral genome such as p17^{gag} would be expected to reflect those in V3. The existence of recombination also provides an explanation for the discordant phylogenies between p17^{gag}, V1/V2, and V3 observed between brain (in the left frontal region) and lymph node sequences in other studies (Donaldson et al, 1994b; Hughes et al, 1997b; Hughes et al, 1997a). While sequences in V3 were tissue specific, sequences in the p17^{gag} and V1/V2 regions were diverse in their relationships between tissues.

Greater degrees of sequence complexity may partly originate from the presence of different infected cell types in a tissue sample. HIV sequences amplified from the choroid plexus of NA234 showed the greatest diversity in the V3 region, containing variants corresponding to those from lymphoid tissue and brain (Fig. 3.4), consistent with the presence of virus from blood-derived cells and brain parenchyma. Several previous studies have implicated the choroid plexus as a potential entry point for HIV into the CNS (Petito et al., 1999; Falangola et al., 1995) and an investigation based upon V3 sequences found viral genotypes corresponding to both brain and spleen variants in this site (Chen et al, 2000), similar to the results presented here. The proximity of these different cell types in the choroid plexus therefore may provide an opportunity for recombination to occur, in addition to its potential role as a site of entry of HIV into the CNS.

The diversity of V3 sequences in different brain regions of the four study subjects was similar to a previous comparison of variants infecting different brain regions (mean pairwise distance between brain regions, 0.021 (Chang et al, 1998), excluding sequences from the left frontal region that were highly divergent in sequence and

failed to group phylogenetically with sequences derived from other regions of the brain in one study subject). In this specific instance the left frontal sequences may have originated from exogenous contamination of the PCR, or corresponded to an epidemiologically unlinked isolate in the case of a mixed infection. In either case, the observed degree of sequence divergence was unlikely to have originated from sequence change over the course of infection within the study subject.

The diploid nature of retroviruses suggests that recombination serves a central function in their replication and evolution. Potential advantages of recombination include the expansion of genetic diversity, and the rescue of replication defective viruses. A large number of investigations using *in vitro* assay systems have established that given the correct conditions, recombination between genetically distinct retroviral isolates is a common event (Clavel et al, 1989; Yu et al, 1998; Hu and Temin, 1990a; Jetzt et al, 2000; Zhuang et al, 2002; Hu and Temin, 1990b). Likewise there is abundant evidence for frequent inter-subtype recombination between quite divergent strains of HIV-1 (Carr et al, 1998; Bobkov et al, 1998; Liitsola et al, 1998; Gao et al, 1998; Liitsola et al, 2000; Montavon et al, 1999), and in many cases this has led to substantial spread of circulating recombinant forms (Gao et al, 1996b; Gao et al, 1998). Evidence for any selective advantage that these frequently observed recombinants might have (for example in allowing more efficient passage through a particular human population) is currently lacking. Thus at a population level recombinant HIV genomes may represent broadly equivalent viral phenotypes, in which case the differing prevalence of particular recombinants would correspond to stochastic events rather than selective advantages. In contrast the indirectly inferred high rate of recombination between HIV-1 variants that have

diverged within a single infected individual as described here and in other studies (Dykes et al, 2000; Brown and Cleland, 1996), and the recent direct demonstration of the same phenomenon (Jung et al, 2002), could have a profound influence on the adaptive potential of HIV-1 variants *in vivo*. Frequent recombination would provide a powerful mechanism for adaptation to several simultaneously acting selective pressures. For example, a prominent feature of the natural course of HIV infection is continuously high levels of virus production and immune-mediated destruction (Ho et al, 1995; Wei et al, 1995; Perelson et al, 1996; Perelson et al, 1997). Selection pressures imposed by cytotoxic T lymphocytes, humoral immune responses, and other immune effector mechanisms may each act upon the products of different HIV subgenomic regions. Reshuffling of genetic information by recombination may allow different collections of mutations that bestow some protection from these selection pressures to come together, thereby providing a selective advantage to the resultant chimeric genome. Similarly, patterns of mutations that confer resistance to different antiretroviral agents are often complex (Jacobsen et al., 1995; Winters et al., 1998; Larder and Kemp, 1989; Richman et al, 1994). Tissue culture infection experiments have shown that recombination can bring together different collections of drug resistance mutations present in separate HIV-1 strains, and thereby produce a new multi-drug resistant virus variant (Moutouh et al., 1996). Frequent *in vivo* recombination such as described here would be expected to allow the occurrence of similar events.

In the specific context of HIV-1 in the brain, recombination between brain and lymphoid virus populations could provide a number of adaptive advantages. It has been suggested that the genetic compartmentalisation often evident upon comparison

of *env* sequence segments between brain and lymphoid-derived HIV-1 viral variants may be a result of adaptation for replication in the unique microenvironment and cell types within the brain (Strizki et al, 1996; Shieh et al, 2000; Power et al, 1994; Power et al, 1998; Gorry et al, 2001; Gorry et al, 2002a). If this is the case then recombination between brain and lymphoid viral populations may provide a simple method for viral variants in the brain to acquire CTL escape mutations or antiviral resistance mutations that evolved in the lymphoid system, while still retaining the genetic determinants of their neuroadapted phenotype. The converse may also be true, with sub-optimal drug concentrations in the CNS potentially providing an ideal environment for the development of drug resistant variants (Lewis et al, 1996; Yarchoan et al, 1988). However, alterations in Env-receptor interactions associated with neuroadaptation may result in a concomitant increase in susceptibility to neutralising antibodies (Gorry et al, 2001; Gorry et al., 2002b; Martin et al, 2001). Therefore recombination between brain and lymphoid-derived HIV-1 variants may allow lymphoid virus to acquire constellations of mutations associated with drug resistance from virus in brain, while still maintaining their lymphoid-adapted phenotype.

Identifying at higher resolution the specific sites and frequencies of recombination events between HIV-1 genomes that have diverged within single individuals requires the sequencing of longer, variable, regions of the viral genome derived from single viruses present *in vivo*. This subject was the topic of further research (Chapter 4). Understanding what contributes to neurotropism will illuminate the selective pressures (if any) that produce the recombinant viruses observed in this study.

Chapter 4: Results (ii)

4.1 Introduction

HIV-1 virions are covered with non-covalently associated heterodimeric glycoprotein spikes comprising a cell-surface attachment protein (gp120), and a membrane spanning fusion protein (gp41) arranged as trimers in the viral envelope. Infection of cells by HIV-1 usually takes place via a sequential series of events beginning with gp120 binding to CD4 on target cells (Maddon et al, 1986; Klatzmann et al, 1984b; Dalglish et al, 1984), leading to a conformational shift in the envelope glycoproteins allowing binding to a second cell surface (coreceptor) molecule (Feng et al, 1996; Alkhatib et al, 1996b; Choe et al, 1996; Deng et al, 1996; Doranz et al, 1996; Dragic et al, 1996). This second interaction exposes domains in gp41 allowing fusion of the viral and cellular membranes followed by entry of the viral core into the target cell. gp120 and gp41 are encoded by the envelope (*env*) gene, consequently this gene contains the principle determinant of receptor usage (Speck et al, 1997; Ross and Cullen, 1998; Chan et al., 1999; Bieniasz et al., 1997), and therefore the differential cellular (Hwang et al, 1991) and potentially tissue tropism displayed by HIV. The infidelity of reverse transcriptase (Preston et al, 1988) coupled with the high level of HIV replication throughout infection (Ho et al, 1995; Wei et al, 1995), and the long duration of infection, make genetic mutations in this region extremely common. These mutations can lead to alterations in viral properties such as receptor usage (Bjorndal et al., 1997), cellular tropism (Schuitemaker et al., 1992) and the susceptibility of the virus to specific immune responses (Wolfs et al, 1990).

Perhaps the most well-documented of these effects are the genetic mutations leading to positively charged residues at specific sites in the V3 loop that are associated with the change in co-receptor usage from CCR5 to CXCR4 (Bjorndal et al, 1997; Speck et al, 1997; Schuitemaker et al, 1992; Chesebro et al., 1992), a phenotypic switch observed during disease progression in approximately 50% of individuals infected with HIV-1 group M subtype B strains. These mutations correlate to some extent with the ability of CXCR4 using (X4) syncytium inducing (SI) primary viral isolates to infect transformed CD4⁺ T cells in culture (T-tropic), and CCR5 using (R5) non-syncytium inducing (NSI) variants to infect macrophages in culture (M-tropic). A wide range of alternative coreceptors can also facilitate HIV-1 entry (at least *in vitro*), and use of these additional coreceptors has proved more difficult to map to specific amino acid residues. Genetic determinants of coreceptor usage or cellular tropism have variously been mapped to the V1/V2 (Cho et al, 1998; Hoffman et al., 1998; Ross and Cullen, 1998; Koito et al, 1994; Boyd et al., 1993; Groenink et al, 1993), V3 (Cann et al., 1992; Hung et al., 1999; Hwang et al, 1991; Kato et al., 1999; Hoffman et al, 1998; Speck et al, 1997), and V4/V5 (Cho et al, 1998) regions of *env*, most frequently in conjunction with specific sequences in other regions of gp120. For example the ability of particular chimeric HIV-1 envelope proteins to mediate CCR3-dependent infection is determined by V1/V2 sequences in conjunction with a CCR5 tropic V3 loop (Ross and Cullen, 1998). Certain mutations in CD4-independent HIV-1 strains generated in tissue culture map to other regions of *env* including C2, C3 (Dumonceaux et al, 1998), and C4 (Labranche et al, 1999), consequently mutations at these or other sites may allow entry of viral variants into CD4-negative cells *in vivo*. A number of other receptors such as DC-SIGN

(Geijtenbeek et al, 2000a), DC-SIGNR (Pohlmann et al, 2001b), Gal-C (Harouse et al, 1991), and heparan sulphate (Patel et al, 1993) may be important for aiding HIV-1 adsorption to cells. Differences in affinity for these molecules could affect cellular tropism by allowing certain HIV-1 variants to preferentially adhere to cell types that express them. For example a higher V3 positive charge provides an enhanced interaction between gp120 and polyanionic molecules such as heparan sulphate (Moulard et al, 2000). Moreover HIV-1 isolates often vary both in their entry receptor density requirements (Kozak et al, 1997; Kabat et al, 1994), and in the precise manner that they interact with coreceptor molecules for entry (Picard et al, 1997; Brelot et al, 1999). Envelope determinants of tropism are therefore complex and may often be conferred by several, discontinuous, regions of the *env* gene.

Target cell factors in addition to the presence of particular viral receptors also appear to influence viral tropism in a complex manner. For example macrophages and microglia express CXCR4 and may be infected by many X4 primary viral isolates but not tissue culture lab adapted X4 viruses (Simmons et al, 1998; Strizki et al, 1996; Ghorpade et al, 1998a; Verani et al., 1998). This may be connected to low levels of CD4 expression on macrophages (Kozak et al, 1997; Platt et al., 1997), or the relatively poor ability of CXCR4 to associate with CD4 on these cells compared with CCR5 (Lapham et al, 1999; Dimitrov et al., 1999). Alternatively, the inefficient replication of many X4 strains in macrophages may be due to post-entry factors, such as the inability of gp120-CXCR4 binding on these cells to activate a requisite signal transduction pathway allowing productive infection (Schmidtayerova et al., 1998). The conformation that coreceptors adopt may also vary in a cell type specific manner (Lee et al, 1999a), such as the formation of CCR5 dimers (RodriguezFrade et al.,

1999) or other cell surface receptor interactions (Mellado et al., 1999). Issues such as this may be especially significant when examining HIV in the brain, where genetically distinct viral populations are frequently observed (Ball et al, 1994; Di Stefano et al, 1996; Donaldson et al, 1994a; Epstein et al, 1991b; Haggerty and Stevenson, 1991; Korber et al, 1994a; Steuler et al, 1992) and several different cell types appear to be infected. These cell types include those that do not express CD4 such as astrocytes (Bagasra et al, 1996; Ranki et al, 1995; Wiley et al, 1986), which leaves open the possibility that individual viruses derived from this organ would display a range of subtly different entry receptor requirements depending upon the cell type they are specific for. Further alterations of viral phenotype in the brain compared to the lymphoid system may be mediated by the microenvironment within this organ. For example neutralising antibodies may be less abundant in the brain than elsewhere in the body, potentially allowing viral envelopes at this site to adopt a more "open" or "pre-triggered" conformation with an increased affinity for coreceptors and a reduced requirement for CD4 (Martin et al, 2001; Gorry et al, 2002a). The presence of fewer immune cells may also lead to lower concentrations of inhibiting factors such as the β -chemokines macrophage inflammatory protein (MIP)-1 α , MIP-1 β , and regulated upon activation normal T cell expressed and secreted (RANTES) as compared with the lymphoid system. Overall, it is clear that the cellular tropism of HIV is dependent upon both the infecting viral strain and target cell/tissue type.

Furthermore at a genetic level the importance of different *in vivo* viral evolutionary mechanisms in the adaptation to new cell and tissue types is still unclear. Recombination clearly has a central role in production of new variants on a global

scale (Robertson et al, 1995; Gao et al, 1996b) yet it has proved difficult to ascertain the specific sites and frequencies of recombination over the course of a single infection. The majority of investigations have indirectly inferred the occurrence of recombination between variants that diverged in a single infected individual (Morris et al, 1999; Dykes et al, 2000; Brown and Cleland, 1996). It would therefore be useful to further characterise this phenomenon in order to assess its possible relevance to escape from antiviral drugs or immune responses, and its potential for aiding viral expansion into new cell/tissue types.

Given these complex interactions, it is of central importance that any genetic or phenotypic study of HIV *env* uses envelope sequences that accurately reflect those present *in vivo*. Currently the preferred method for amplifying genes from HIV is the polymerase chain reaction (PCR), as this does not lead to the immediate and intense selection of small subsets of virus associated with cell culture systems (Sanchezpalomino et al., 1993). However, PCR is also known to produce *in vitro* artefacts, most significantly the artefactual recombination of parent strands during amplification (Meyerhans et al., 1990; Yang et al., 1996; Zylstra et al., 1998; Judo et al., 1998; Fang et al., 1998) and the incorporation of single nucleotide errors by the DNA polymerase used in the reaction (Saiki et al., 1988; Bracho et al., 1998; Smith et al, 1997; Bracho et al, 1998). Making every effort to avoid the resulting potential for scrambling of hypervariable regions and incorporation of point mutations is critical to any study attempting to identify discrete genetic or phenotypic properties of specific viral populations. Moreover, molecular cloning is often used subsequently to PCR to allow sequencing and expression of amplified genes yet this method may also lead to artefactual results and erroneous interpretations. If small numbers of

template molecules are present in the beginning of a PCR reaction, or if preferential primer binding has led to the disproportionate amplification of certain templates, then several clones may be derived from a single virus or provirus (Liu et al., 1996). This procedure can therefore result in misrepresentation of the genetic diversity within a sample. Considering the importance of *env* in understanding many areas of HIV pathogenesis, and the large volume of research focussing on this gene, we aimed to quantify each of these artefactual effects and if possible reduce or eliminate them thus providing genuine *in vivo* HIV-1 *env* sequences for analysis.

4.2 Aims

- 1) Individually assess the significance of a range of *in vitro* artefacts associated with the PCR amplification and subsequent cloning of the HIV-1 *env* gene.
- 2) Develop methods for eliminating these artefacts and generating sequences that are representative of those present *in vivo*.
- 3) Further previous observations of *in vivo* recombination by analysing *env* sequences obtained at limiting dilution from separate tissues.

4.3 Materials and methods

As described in Chapter 2.

4.4 Results

4.4.1 Nucleotide misincorporation during PCR

One problem associated with use of PCR is the introduction of mutations resulting from nucleotide misincorporations during amplification. We therefore investigated the extent to which this occurred in a nested PCR encompassing the entire *env* gene of HIV-1 along with the first 300 bp of *nef* using previously described primers (Gao et al., 1996a). 30 copies of either full-length HIV-1_{LAI} or HIV-1_{CAMI} (of known sequence) were amplified in three separate *env* PCR reactions and a total of seven clones derived from these reactions were fully sequenced (3 clones each from 2 PCR products and 1 clone from the 3rd PCR product). An average of 1.57 mutations/clone was identified (11 mutations in 7 clones) giving an overall frequency of nucleotide misincorporation in final product clones of 1 per 1839 bp. Calculating the error rate for a polymerase requires the real number of DNA duplications during PCR to be known, as opposed to the number of thermal cycles the DNA is subjected to (70 in total for this nested PCR). Given an average weight of 660 g/mole per base-pair for double stranded DNA, one mole of *env* (2887 bp) has a mass of 1.9×10^6 g, therefore a single molecule of *env* has a mass of 3.16 ag ($1.9 \times 10^6 / 6.022 \times 10^{23} = 3.16 \times 10^{-18}$ g). Assuming an input number of 30 molecules and 5 μ g of end product DNA, then the total amplification achieved by PCR is approximately 5×10^{10} (i.e. $5 \times 10^{-6} / (30 \times 3.16 \times 10^{-18}) = 5.27 \times 10^{10}$), corresponding to approximately 36 duplications of DNA (5×10^{10} is approximately 2^{36}). 1.57 errors per 1×10^5 incorporations (2887×36) gives an error rate for Expand High Fidelity polymerase (a mixture of *Pwo* and *Taq*) in this PCR of 1.51×10^{-5} , lower than that reported for other thermostable polymerases (Smith et al, 1997) but still potentially harmful for studies of the

amplified clones. For example 7/11 of the polymerase misincorporations identified in this study resulted in predicted amino acid changes (Fig. 4.1). These included a mutation within the V2 region of *env* (position 552) and the replacement of a conserved cysteine residue essential for maintaining the di-sulphide bond of the V3 loop with a tyrosine residue (position 1075). Mutations of this type are likely to result in translated glycoproteins that display altered biological properties from the original templates and may also produce non-functional proteins.

4.4.2 *In vitro* recombination during amplification by PCR

A second problem associated with the PCR amplification of genetically distinct templates is the production of artefactual recombinant molecules. Therefore a study was undertaken to determine if entire *env* gene PCR amplification of heterogeneous HIV templates leads to significant *in vitro* recombination and final products that are chimaeras of the original sequences. A system was devised to assess the frequency of recombination *in vitro* using the two genetically distinct HIV-1 template clones HIV_{LAI} and HIV_{CAM1}. These contain unique *Hind* III and *Xmn* I restriction sites respectively, spaced 1285 bp apart (Fig. 4.2.A).

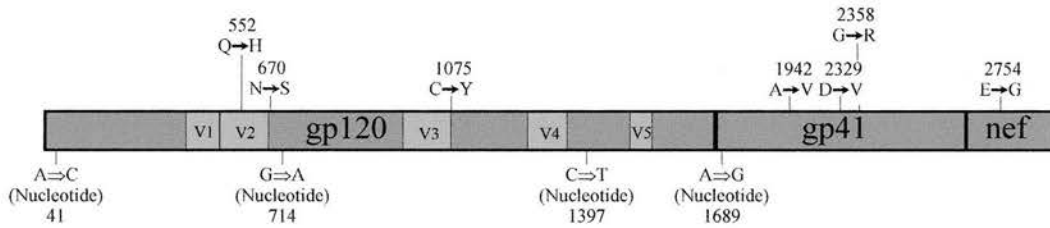


Figure 4.1. Schematic representation of mutations that occurred during *env* PCR using cloned viral templates. A summary of mutations identified in seven *env* clones derived from 3 PCR reactions using cloned HIV-1 genomes as template. These mutations are due to misincorporations by the Expand High Fidelity PCR System (a mixture of *Pwo* and *Taq*) used for amplification. Amino acid changes due to non-synonymous changes are shown above and synonymous nucleotide changes below the *env* diagram. Numbering is from the start of *env* and refers to nucleotide position.

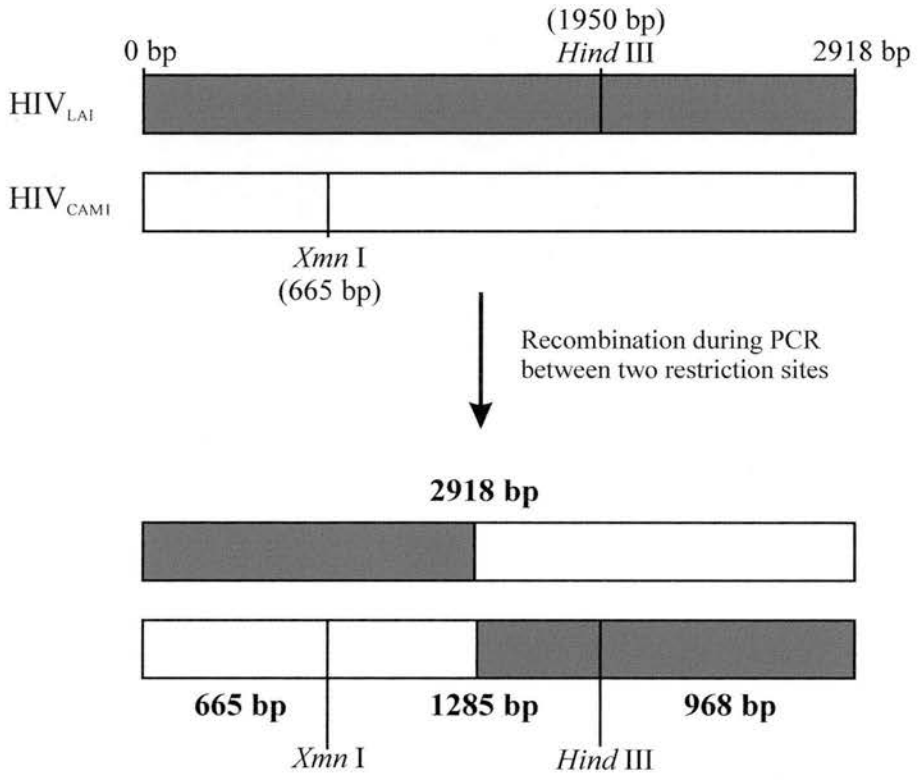
Figure 4.2. System for detecting recombination during *env* PCR amplification.

A) Positions of *Hind* III and *Xmn* I restriction sites in *env* PCR amplicons derived from HIV_{LAI} or HIV_{CAMI}. If the two genotypes are mixed and the *env* gene amplified by PCR then no recombination would lead to the presence of only parental genotypes in the final product. Recombination during PCR between the two restriction sites spaced 1285 bp apart would produce sequences with neither restriction site (5' end HIV_{LAI}, 3' end HIV_{CAMI}), leading to an uncut band of 2918 bp upon *Hind* III/*Xmn* I digestion, or both restriction sites (5' end HIV_{CAMI}, 3' end HIV_{LAI}), producing bands of 665 bp, 1285 bp, and 968 bp upon digestion. The ratio of the four products revealed by enzymatic cleavage indicates the frequency of recombination during PCR.

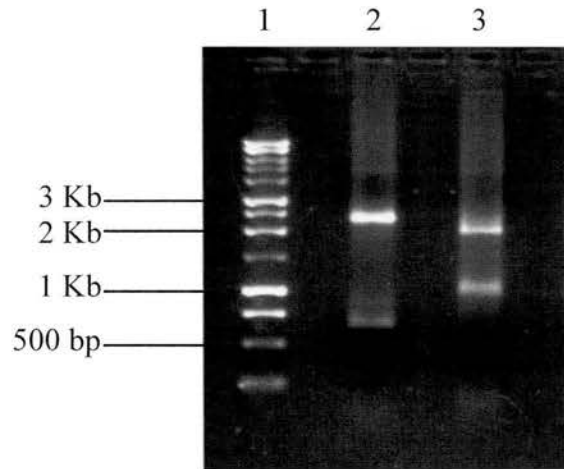
B) *Hind* III and *Xmn* I double restriction digest on *env* PCR products derived from only HIV_{CAMI} (lane 2), or HIV_{LAI} (lane 3) demonstrating the expected restriction pattern for each, a 1 Kb DNA ladder is also shown (Lane 1).

A)

env PCR product



B)



Therefore if the two clones are mixed before amplification of the *env* gene by PCR, products generated by an *in vitro* recombination event between the two sites can be easily identified by possession of either both or neither restriction site. *Hind* III and *Xmn* I double digests on *env* PCR products derived from only HIV_{CAMI} or HIV_{LAI} templates demonstrated the expected restriction pattern for each (Fig. 4.2.B). It has been previously suggested that *in vitro* recombination may occur more frequently where large numbers of target DNA molecules are initially present in the reaction (Meyerhans et al, 1990). Hence, equimolar mixtures containing 3000, 300 or 30 copies of each clone as template were prepared and amplified by *env* PCR. HIV-negative peripheral blood mononuclear cell DNA was also added to the reactions prior to amplification in order to more closely simulate amplification of proviral *env* genes from DNA extracted from infected cells. These products were cleaved with *Hind* III and *Xmn* I in a double restriction digest and the bands sized by agarose gel electrophoresis (Fig. 4.3.A). The detection of visible recombinant bands of approximately 2.9 Kb and 1285 bp corresponding to sequences containing neither or both of the restriction sites indicated that *in vitro* recombination had occurred at a relatively high frequency between the two restriction sites on the parent DNA strands. Note that the 2.9 Kb uncut band appears more intense than the 1285 bp doubly cleaved band. This is likely to be due to the re-annealing of fully extended products competing significantly with primer binding in the latter stages of the PCR, leading to the formation of heterodimeric molecules that are refractory to restriction digestion. Therefore the presence of the 1285 bp restriction fragment is the more rigorous indication of recombination.

Figure 4.3. Detection of recombinant *env* PCR products.

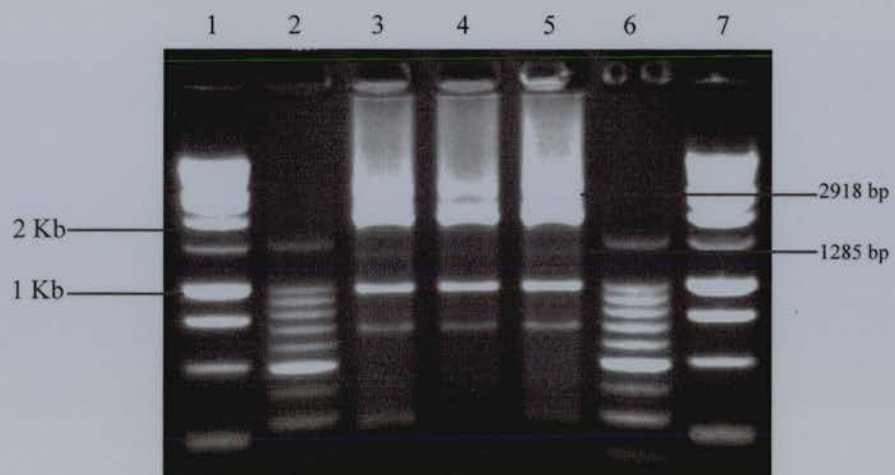
Recombinants were identified by restriction mapping after amplification of different concentrations of the HIV_{CAMI}/HIV_{LAI} genotypic mix. The levels of the 2918 bp and 1285 bp bands (the presence of which demonstrates that recombination has taken place) are also indicated.

A) Lanes 1 and 7: 1 Kb DNA ladder, Lanes 2 and 6: 100 bp DNA ladder, Lanes 3, 4, and 5 represent *Hind* III/*Xmn* I restriction digest products of PCR reactions that contained 3000, 300, or 30 copies respectively of each clone as template. Bands at 2.9 Kb and 1285 bp are evident in each of the digested reaction products.

B) Digests of PCR products amplified from the HIV_{CAMI}/HIV_{LAI} mix after dilution to near PCR end point. Lanes 1 and 13: 1 Kb DNA ladder, Lanes 2 and 12: 100 bp DNA ladder, Lanes 3-11: digests of replicate PCR reactions (see text for information on the final composition of each of these reaction products).

C) Digests of replicate PCR products amplified from a further 1 in 3 dilution of template shown in B. Lane 1: 1 Kb ladder, Lanes 2-10 replicate PCR products digested with both *Hind* III and *Xmn* I (see text for further details of product composition).

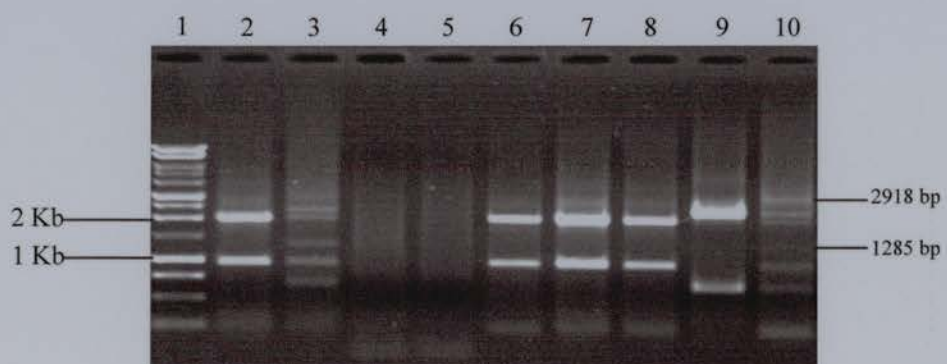
A)



B)



C)



To assess the artefactual recombination frequency when low numbers of template molecules are amplified by PCR this genotypic mixture was then diluted out to near PCR end point and PCR replicates were performed at the last positive (10-fold) dilution. Reaction products were again digested with both *Hind* III and *Xmn* I to characterise their genetic composition (Fig. 4.3.B). Lane 5 represents a negative PCR result, as would be expected when approaching limiting dilution. Assuming a Poisson distribution it was calculated that each replicate contained an average of approximately 2.2 sequences from the original HIV_{CAMI}/HIV_{LAI} mix. Reactions 4 and 7 contained only HIV_{CAMI} and HIV_{LAI} respectively and the remaining 5 reactions contained genotypic mixes of the two original templates and displayed varying intensities of the 1285 bp recombination indicator band. The genotypic mix was then further diluted 1 in 3 and figure 4.3.C shows digested products derived from this dilution, which yielded 7 positive reactions from 9 tested. This resulted in varying outcomes: 4 isolated HIV_{LAI} sequences (lanes 2, 6, 7, and 8), 1 of HIV_{CAMI} (lane 9), 2 HIV_{CAMI}/HIV_{LAI} recombinant patterns (lanes 3 and 10) and 2 negative replicates (lanes 4 and 5). The replicate in lane 3 containing both HIV_{LAI} and HIV_{CAMI} sequences produced approximately equimolar amounts of the 4 possible restriction patterns and the replicate in lane 10 displayed a less intense recombinant pattern than that indicative of the original sequences present in the reaction. These results implied a loss of genetic linkage between the restriction sites 1285 bp apart under these PCR conditions, and demonstrated that recombination in PCR can also occur at high frequency with low numbers of input molecules. All PCR replicates performed on more concentrated dilutions than in figure 4.3.C demonstrated the presence of both genotypes along with varying intensities of recombinant bands. Any PCR positives

observed at lower dilutions than that in figure 4.3.C demonstrated the presence of only a single genotype by restriction analysis (data not shown). The isolation of single genotypes in certain PCR replicates also suggested that the sensitivity of the *env* PCR is sufficient for the amplification of single proviral molecules at limiting dilution from an original genotypic mix, a procedure permitting the elimination of artefactual recombination

4.4.3 Limiting dilution *env* PCR

To confirm that the limiting dilution technique could be applied to the *env* PCR, the sensitivity of this reaction was compared with that of a PCR using primers spanning the V3 region, which had previously proven sensitive enough for use on single proviral templates (Donaldson et al, 1994a). Cloned HIV_{LAI} DNA was titrated in a 3-fold dilution series and whole *env* and V3 PCRs were both performed in triplicate reactions on each of the dilutions in the series (Table 4.1). The results indicated that with cloned DNA as template the whole *env* PCR had similar sensitivity to that of V3. An additional sensitivity comparison between the V3 and *env* PCRs was then undertaken using DNA extracted from an autopsy specimen recovered from the brain of an individual with giant cell encephalitis (right occipital lobe of NA021, Chapter 3). Serial 10-fold dilutions of the DNA were analysed by both V3 and *env* PCR, and replicate PCR reactions using both sets of primers were then performed at limiting dilution (Table 4.2). The sensitivity of whole *env* PCR was similar to PCR using V3 primers although there were approximately five-fold fewer positive end point replicates using the *env* PCR.

PCR method	Number of positive replicates/number tested						
	*300	100	30	10	3	1	0.3
V3 primers	3/3	3/3	3/3	3/3	1/3	1/3	0/3
<i>env</i> primers	3/3	3/3	3/3	2/3	2/3	1/3	0/3

Table 4.1. Sensitivity comparison of *env* PCR with V3 PCR using cloned HIV proviral DNA as template. *Approximate number of DNA copies added to PCR.

PCR method	DNA dilution			
	Undiluted	10 ⁻¹	10 ⁻²	10 ^{-3*}
V3 primers	+	+	+	14/30
<i>env</i> primers	+	+	+	3/30

Table 4.2. Sensitivity comparison of *env* PCR with V3 PCR using HIV proviral DNA extracted from autopsy tissue as template. *30 replicate reactions carried out at end point dilution for each PCR with frequency of positives indicated.

The majority of DNA extracted from autopsy tissue was present as a high molecular weight (> 10 Kb) species (data not shown) therefore autolysis or cleavage of template DNA during extraction is unlikely to account for this small difference in sensitivity. It is possible that less of the 3 Kb *env* template was present in a form suitable for PCR amplification compared with the shorter 350 bp V3 template.

4.4.4 Genetic diversity comparison between limiting dilution and cloning

Resampling error occurs when low input copy numbers or preferential primer binding during PCR results in cloned sequences that significantly under-represent the genetic diversity within an original DNA sample. Resampling due to low input copy number is related to the number of templates added to a PCR reaction and the number of clones derived from that reaction, allowing the probability of resampling at least one original template to be calculated (Liu et al, 1996). For example if 25 input templates are added to a PCR reaction, and 10 clones derived from that reaction are sequenced then there is an 88% probability of resampling at least one of the original templates (Liu et al, 1996). This effect can be reduced by first quantifying the amount of template in a sample then ensuring that a large number of template molecules are added to the PCR. However, the issue of preferential primer binding cannot be effectively accounted for by varying template copy number. Furthermore, the exponential amplification involved in PCR means that even small differences in primer affinity to template DNA can result in a minority sequence becoming the most abundant genotype at the end of a PCR reaction.

To specifically address this issue 9 *env* sequences were amplified at limiting dilution from the right occipital lobe of patient NA021 as described above and the V1/V2

region sequenced. In parallel the V1/V2 region was directly amplified by PCR from the same DNA sample in a reaction containing over 1000 original HIV-1 proviral DNA templates using *Taq* DNA polymerase. *Taq* polymerase was used in order to more closely imitate the majority of similar investigations using PCR and cloning techniques reported in the literature. The single V1/V2 PCR product was cloned and 11 individual clones were isolated and fully sequenced. Given an input DNA copy number of over 1000 with 11 clones sampled, the average number of clones that should be derived from unique templates is over 10.9 and the probability of one or more resampling events occurring is less than 0.054, or 5.4% (Liu et al, 1996). Therefore resampling of clones is unlikely to be evident and the genetic diversity observed within the cloned V1/V2 sequences should be very similar to that observed in limiting dilution PCR products. Note that sporadic polymerase errors are more likely to have occurred in the V1/V2 PCR than the V1/V2 region of an *env* PCR product. This is due to the lower fidelity *Taq* polymerase used for V1/V2 amplification, and to the greater number of polymerase incorporations that will have occurred during generation of the V1/V2 final PCR product compared with the V1/V2 region of the longer *env* PCR product. The previously determined number of misincorporations during *env* PCR leads to an expected 1 polymerase error in the entire data set of V1/V2 sequences from limiting dilution PCR products, a figure that is likely to be greater in the directly amplified V1/V2 PCR products described here. N-J phylogenetic trees using J-C distances were created with each data set using HIV_{MN} as an outgroup (Fig. 4.4).

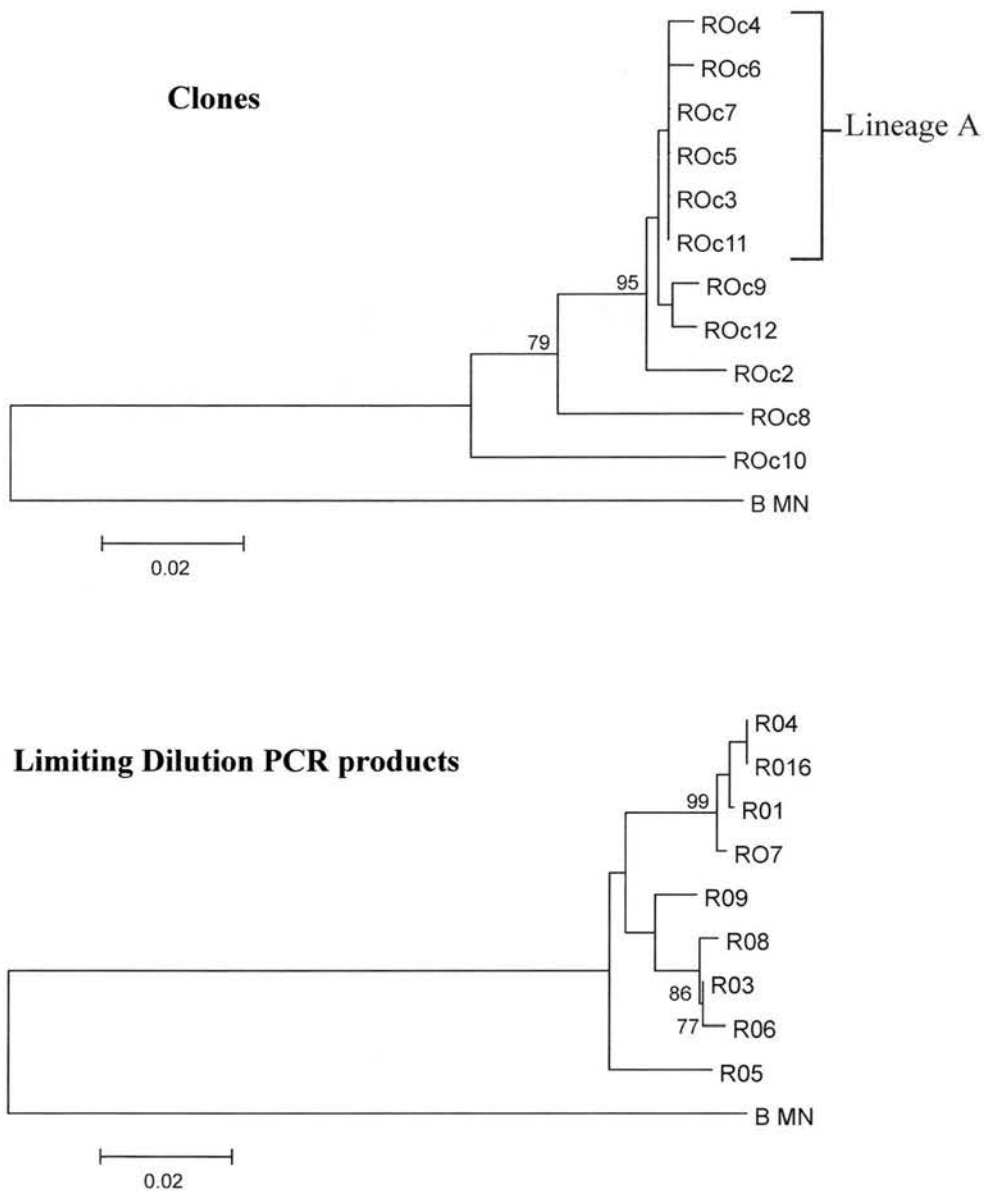


Figure 4.4. Phylogenetic trees constructed with V1/V2 sequences obtained by cloning or limiting dilution PCR. Upper panel: Cloned sequences, Lower panel: PCR product sequences. Divergence between nucleotide sequences was estimated using Jukes Cantor distances (scales indicated below trees), and trees were constructed from the distance matrices by the neighbour-joining method. The robustness of observed groupings was tested by bootstrap resampling of 500 data sets with values over 75% indicated on branches. The subtype B isolate HIV_{MN} was used as an outgroup to root the tree. Lineage A corresponds to the set of closely related cloned sequences referred to in the text.

The tree produced from cloned sequences appeared less diverse than that from limiting dilution PCR products. This impression was largely due to one lineage (Fig. 4.4, Lineage A) displaying a marked reduction in genetic diversity compared with the topology of the rest of the tree and that of limiting dilution PCR sequences. The sequences within this lineage were remarkably similar and contributed to a median number of nucleotide differences compared with the 60% majority sequence (ignoring insertion/deletion events) per sequence of 1 for cloned sequences in contrast to 7 for limiting dilution PCR products (Fig. 4.5). The number of differences compared to the consensus was a somewhat arbitrary value because if resampling had occurred then these duplicated sequences would also contribute towards the composition of the consensus sequence. However, other possible measures such as comparing the mean genetic distances within each group of sequences were not appropriate as the presence of a few divergent “outlier” strains would potentially compensate for other, more similar sets of sequences elsewhere in the data set. Clones ROc7 and ROc11 each had an identical change that was shared with sequences from limiting dilution PCR products (position 151). These two cloned sequences were therefore probably not derived from the same template as the remainder of sequences in Lineage A, although it is possible that they were derived from the same sequence as each other. Clones ROc3, ROc4, ROc5, and ROc6 contained only two nucleotide mutations in total. These mutations (position 60 in ROc4, and position 181 in ROc6) were unique within the data set and unprecedented within the Los Alamos National Laboratory HIV sequence database. Hence these two mutations probably represented polymerase misincorporations.

Given that the original proviral templates were fairly diverse in V1/V2, as demonstrated by the "correct" sequences from limiting dilution PCR products, the four cloned sequences: ROc3, ROc4, ROc5, and ROc6 are most likely derived from a single original template.

Significantly this result did not simply reflect a lack of original template molecules in the PCR. The same V1/V2 PCR used prior to cloning was also used to quantify the original number of amplifiable templates, an approach that is essential for adequate template numbers to be assured (Rodrigo et al., 1997). PCR amplification of DNA containing over 1000 of these amplifiable template molecules with 11 clones sequenced is unlikely to have produced a single resampling event because of lack of input DNA. Therefore the resampling of clones documented here likely reflects preferential primer binding or random events during the PCR leading to an over representation of certain genotypes in the final product. Such events are difficult to account for completely, even when primers are designed to correspond to fairly conserved regions of the HIV genome.

4.4.5 Sequence Analysis of limiting dilution *env* sequences

Given that the sensitivity of the *env* PCR had proven sufficient for amplification from single HIV-1 proviral templates, this approach was used to amplify and directly sequence the envelope genes from two regions of brain and one lymph node obtained at autopsy from an HIV infected individual displaying giant cell encephalitis (study subject NA021, Chapter 3). The right occipital and left temporal lobes were chosen for this analysis because they had previously been shown to contain virus genetically distinct or similar respectively to lymph node virus in p17^{gag} yet each were

completely segregated from lymph node in the V3 region (Chapter 3). Therefore upon closer examination of the *env* gene perhaps would display unpredictable interrelationships and further evidence of recombination between virus present in different tissues. A total of 9 sequences were obtained from the right occipital lobe (RO), 8 from left temporal (LT), and 8 from lymph node (LN, or LN2 as described in Chapter 3). An example of several final *env* PCR products obtained at limiting dilution is shown (Fig. 4.6). Each product formed a discrete band of approximately 3 Kb, demonstrating the specificity obtained using this PCR. It is also noteworthy that while the Poisson formula may be used to predict the average number of molecules per reaction, it is still possible for positive reactions that appear to be at limiting dilution to contain more than one proviral molecule. Two of the PCR products originally generated displayed completely ambiguous sequences across the V1/V2 and V4/V5 hypervariable regions, probably due to the presence of two or more original templates containing length polymorphisms in these genomic regions. These two sequences were accordingly excluded from the current analysis. Remaining PCR products did not display any such length polymorphisms and were unambiguous during individual sequencing, thereby providing additional confidence that they were indeed derived from single original proviral copies.

The predicted amino acid sequence of the isolated *env* (gp160) genes is shown (Fig. 4.7) with relevant domains annotated.

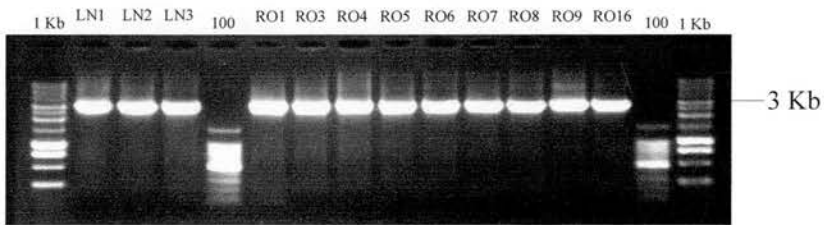


Figure 4.6. Example of *env* PCR products amplified from single proviral templates. Template DNA was extracted from either lymph node (LN) or right occipital lobe (RO) tissue samples. 1 Kb and 100 indicate 1 Kb and 100 bp DNA ladders respectively, level of 3 Kb band is marked.

Figure 4.7. Predicted Env (gp160) amino acid sequences of PCR products amplified at limiting dilution from different tissue samples obtained at autopsy. A 60% majority consensus sequence is included for comparison. Dots indicate residues identical to reference sequence, dashes indicate alignment gaps, "x" indicates frameshift mutations, and stars indicates stop codons. RO: right occipital lobe, LT: left temporal lobe, and LN: lymph node. Hypervariable regions and other important structural features are labelled. Red asterisks indicate sites for CD4 binding and blue asterisks indicate sites where substitutions are critical for CCR5 binding. These binding sites are taken from a previously published alignment (Yamaguchi-Kabata and Gojobori, 2000) and are based upon the X-ray crystal structure of a gp120-CD4 complex (Kwong et al., 1998), and mutational analysis of specific amino acid residues in gp120 (Rizzuto et al., 1998) respectively. Green triangles indicate sites in the V3 region where the occurrence of positively charged amino acids strongly correlates with CXCR4 usage. Figure is continued over four pages.

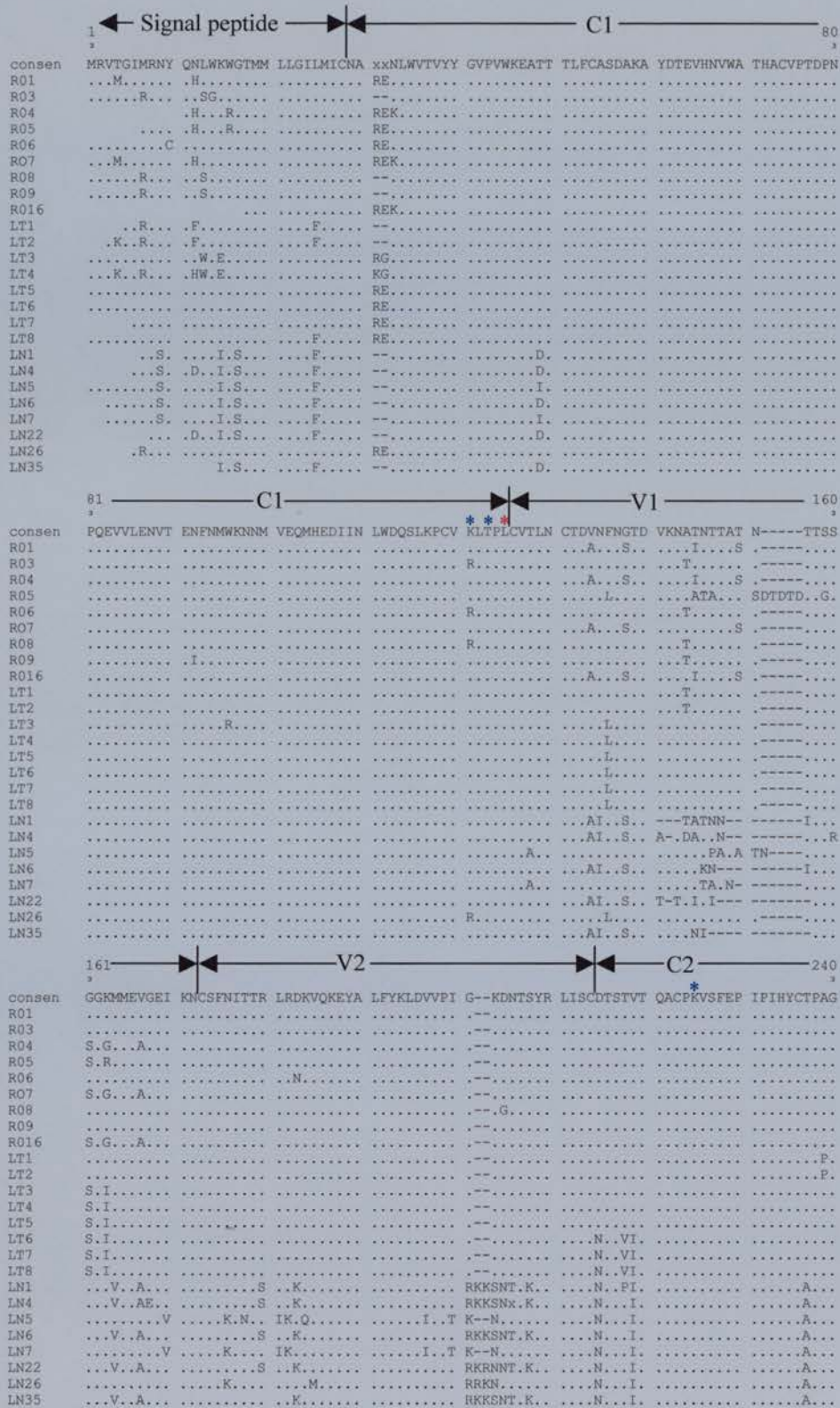


Figure 4.7.



Figure 4.7. Continued

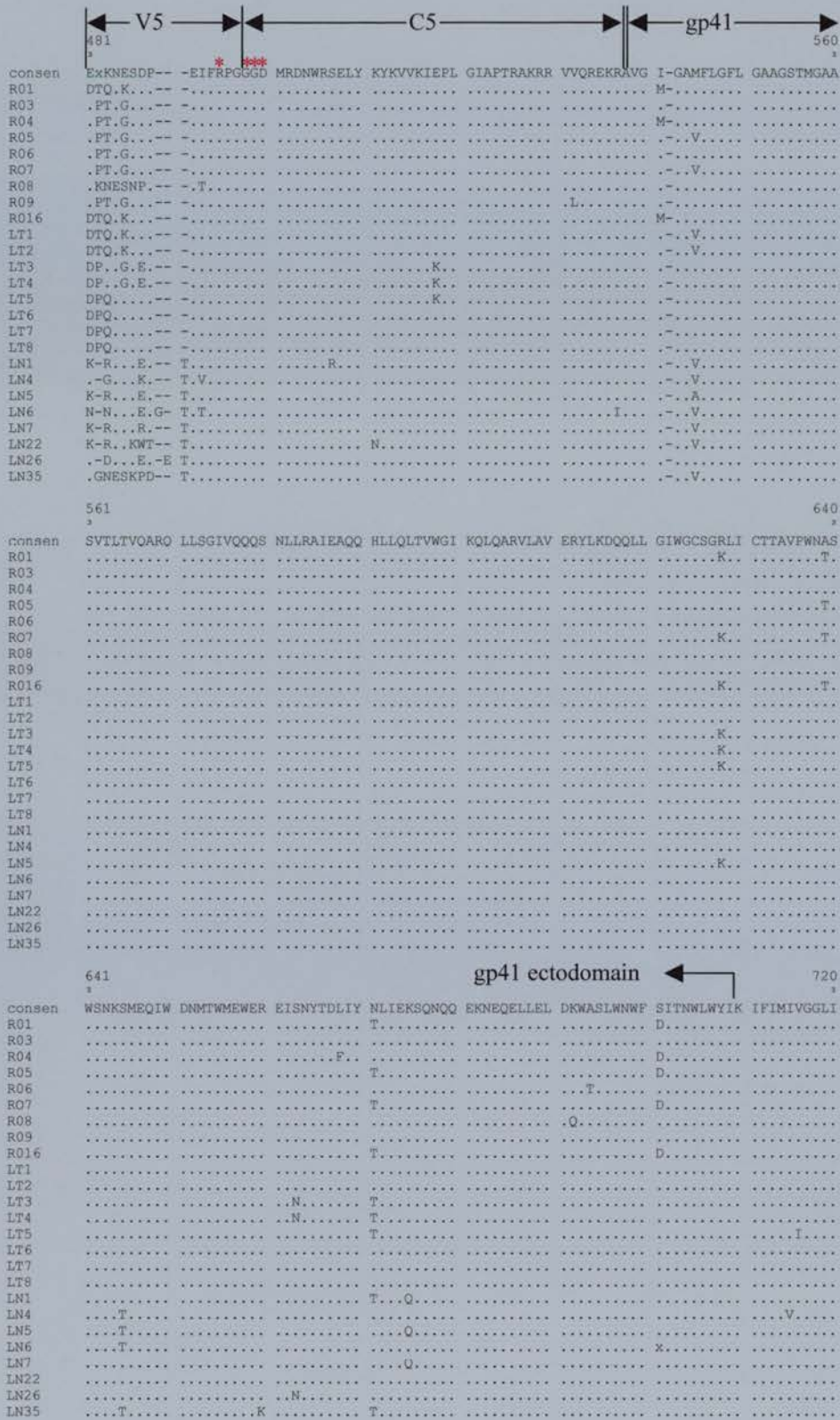


Figure 4.7. Continued

Membrane spanning domain Endocytosis Motif (YXXL) gp41 cytoplasmic domain

721 800

consen	GLRIVFTVLS	IVNRRVQGY	PLSFQTRLPA	RRGPDREGI	EEEGGERDRD	RSGPLVNGFL	AIIVVDLRSL	FLFSYHRLRD
R01F..	P.....S.....
R03S.....
R04I.....S.....
R05F..	P.....S.....
R06
R07F..	P.....S.....
R08S.....
R09
R016F..	P.....S.....
LT1T.....
LT2T.....
LT3F..	P.....S.....
LT4F..	P.....S.....
LT5F..	P.....
LT6F..	P.....
LT7
LT8
LN1	T.T.....
LN4	T.T.....
LN5S.....	T.A.....
LN6	T.T.....N.....
LN7	T.T.....N.....
LN22	T.T.....
LN26A.....
LN35	T.T.....N.....

801 880

consen	LLLIVTRIVE	LLGRR-GWEA	LKYWNLLQY	WSQELKNSAV	SLLNATAIAV	AEGTDRVIEV	VQRAYRAILH	IPTRIRQGLE
R01	L.....
R03	V.....	I.....	I.....
R04	L.....
R05	V.....	I.....	I.....
R06	V.....	I.....	I.....
R07	I.....	I.....
R08	V.....	I.....	I.....
R09	L.....
R016A.....	L.....K.....
LT1	V.....	I.....	I.....
LT2A.....	V.....	I.....	I.....
LT3	L.....
LT4	L.....
LT5	L.....
LT6	V.....	I.....	I.....
LT7	V.....	I.....	I.....
LT8	V.....	I.....	I.....
LN1L.....T..V.....	I.....	I.....
LN4T.....F.....	I.....
LN5T.....L.....	I.....	F.....
LN6*L.....F.....	I.....
LN7T.....L.....F.....	F.....
LN22T.....L.....F.....	F.....
LN26F.....	I.....
LN35L.....F.....	I.....

881

consen	RALL*
R01*
R03*
R04*
R05*
R06*
R07*
R08*
R09*
R016*
LT1*
LT2*
LT3*
LT4*
LT5*
LT6*
LT7*
LT8*
LN1*
LN4*
LN5*
LN6*
LN7*
LN22*
LN26*
LN35*

Figure 4.7. Continued

No two sequences were entirely identical although some, most notably from the LT region were closely related to each other. There also appeared to be a broad segregation between the lymph node and brain-derived sequences. One sequence (LN6) contained a stop codon in the cytoplasmic domain of gp41 (position 810) predicted to cause a premature truncation 74 amino acids from the C-terminus. A single nucleotide deletion (frameshift mutation) was present in both LN4 (position 206) and RO7 (position 770), but the remainder of the open reading frames appeared intact. As expected, the gp120 coding region was more variable than gp41, and the greatest genetic diversity was present within the hypervariable regions. In the V1/V2 region lymph node sequences contained a variety of length polymorphisms whereas only one brain-derived sequence (RO5) had a relative insertion event (Position 152-156). The V3 region was relatively homogeneous with no mutations completely segregating between the separate tissues, although one relative N→V/A change present in the 5' V3 flanking sequence (position 301) did distinguish between LN and brain populations. Sites 324 and 340 in the V3 region are positions where the presence of a positively charged amino acid is strongly indicative of CXCR4 usage. Both of these sites contained uncharged or acidic amino acids, thus all of the Envs in this data set are derived from viruses that would be expected to use CCR5. The limiting dilution V3 sequences described here also correspond well to the consensus V3 sequences described in Chapter 3 that were obtained from the same tissues. In V4 the RO sequences were most variable and two sites consistently segregated between LN and brain-derived sequences (positions 415 and 416). In V5, sequences were also diverse but appeared to group broadly with others from the same tissue. One sequence (LN6) had a relative R→I mutation in the fourth amino acid of the usually

highly conserved (R/L-X-K/R-R) gp120/gp41 cleavage site (position 534-537), and thus may represent a biologically unviable virus. The beginning of gp41 was much more conserved than the majority of gp120, perhaps in part as a result of purifying selection at the nucleotide level mediated by the presence of the Rev response element (RRE) at this point. Other areas of specific functional importance were also relatively conserved. Three sequences contained single sporadic mutations in the gp41 membrane-spanning domain (position 710-733), and the Y-X-X-L endocytosis motif (position 739-742) was a region of complete identity among all sequences.

One way of detecting potentially biologically important differences in brain or lymphoid sequences is to identify those sites known to be important for receptor binding or have been specifically shown to alter receptor usage in laboratory experiments. If such sites are significant for tissue tropism then there may be segregating or mostly segregating mutations at these positions separating brain and lymph node derived sequences. Some of the sites important for CD4 and CCR5 binding as identified by X-ray crystal structure (Kwong et al, 1998) or site directed mutagenesis experiments (Rizzuto et al., 1998) are also labelled on the Env amino acid alignment. These sites are mainly conserved within all sequences, probably reflecting the importance of gp120-receptor interactions. One site important for CD4 binding (position 301) did segregate between brain and lymph node sequences, with an asparagine residue present in all brain derived sequences but no lymph node sequences, and may therefore represent an adaptive mutation. Several sequences contained an arginine instead of lysine at position 121, an important site for CCR5 binding, but not in a tissue specific manner, this difference is therefore unlikely to represent an important adaptive mutation. Alteration of N-linked glycosylation

signals may also alter the properties of translated glycoproteins. One report documenting the spontaneous adaptation in tissue culture of an HIV-1 CCR5 using variant (isolate ADA) for replication in CD4-negative cells describes the movement or loss of an N-linked glycosylation site as a necessary and sufficient factor in the altered viral phenotype (Kolchinsky et al, 1999). This mutation in the V1/V2 stem (*env* amino acid 197 on the HXB2 genome or position 215 in this study) also provided enhanced CD4-independent infection in conjunction with mutations elsewhere in ADA *env*. These alterations, when introduced into YU2 or the CXCR4-using ADA (MMM) glycoproteins did not generate a similar CD4-independent phenotype, suggesting that the effect was restricted to ADA Env. However, in the current study all lymph node derived sequences contained this N-linked glycosylation site and all but three brain-derived sequences did not, due to a relative N→D substitution (amino acid position 215). Other changes noted in the prior study were not present in this data set but the naturally occurring loss of this glycosylation site may to some extent mirror CD4-independence in cell culture isolates by enhancing direct gp120-CCR5 binding and negating or more likely reducing the requirement for CD4 in these interactions. Such alterations in brain-derived isolates may be expected given the apparent infection of cell types that do not express CD4, or express only low levels of this molecule.

A) Tat exon 2

```

1                               31
3
consen  RPASQPGGDP TGPKESKKKV ERETETDPVH *
R01     .....R... ..Q..... ERETETDPVH *
R03     .....R... ..Q..... ERETETDPVH *
R04     .....R... ..Q..... ERETETDPVH *
R05     .....R... ..Q..... ERETETDPVH *
R06     .....R... ..Q..... ERETETDPVH *
R07     .....R... ..Q..... ERETETDPVH *
R08     .....R... ..Q..... ERETETDPVH *
R09     .....R... ..Q..... ERETETDPVH *
R016    .....R... ..Q..... ERETETDPVH *
LT1     .....R... ..Q..... ERETETDPVH *
LT2     .....R... ..Q..... ERETETDPVH *
LT3     .....R... ..Q..... ERETETDPVH *
LT4     .....R... ..Q..... ERETETDPVH *
LT5     .....R... ..Q..... ERETETDPVH *
LT6     .....R... ..Q..... ERETETDPVH *
LT7     .....R... ..Q..... ERETETDPVH *
LT8     .....R... ..Q..... ERETETDPVH *
LN1     .....R... ..Q..... ERETETDPVH *
LN4     .....R... ..Q..... ERETETDPVH *
LN5     .....R... ..Q..... ERETETDPVH *
LN6     .....R... ..Q..... ERETETDPVH *
LN7     .....R... ..Q..... ERETETDPVH *
LN22    .....R... ..Q..... ERETETDPVH *
LN26    .....R... ..Q..... ERETETDPVH *
LN35    .....R... ..Q..... ERETETDPVH *

```

B) Rev Exon 2

```

                                RNA binding
                                and Nuclear
                                Localisation      Oligomerization      Activation
                                Domain
□
1                               91
3
consen  PPSQEGTRQ ARRNRNRWR ERQRQIRSI EWILSNHLGR PAEPVPLQLP FLERLTLDCN EDCGTSGTQG VGSPQILVES PTVLESGTKE *
R01     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
R03     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
R04     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
R05     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
R06     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
R07     .L.P..... ..x..F..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
R08     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
R09     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
R016    .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
LT1     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
LT2     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
LT3     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
LT4     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
LT5     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
LT6     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
LT7     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
LT8     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
LN1     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
LN4     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
LN5     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
LN6     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
LN7     .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
LN22    .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
LN26    .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *
LN35    .L.P..... ..F..... K..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... ..V..... *

```

Figure 4.8. Predicted amino acid sequences of the second exons of Tat and Rev, encoded within the gp160 open reading frame. A) Tat exon 2. B) Rev exon 2. Sequence labels are as for Figure 4.7. The "x" in each RO7 sequence indicates a frameshift that is present in all three overlapping open reading frames. Several important functional domains are also annotated in the Rev alignment.

The second exons of Tat and Rev are present in overlapping open reading frames towards the 3' end the gp41 coding sequence. As selection pressure on *tat* or *rev* sequences may directly alter the *env* sequence, predicted amino acid sequences of the second exons of Tat and Rev are shown (Fig. 4.8). The amino acid sequence of Tat exon 2 shows no complete tissue segregation, although positions 27 and 29 near the C terminus do show a separation of brain from lymph node sequences with the exception of LN26. Of interest in Rev exon 2 is the almost complete conservation of the putative nuclear localisation and RNA binding domain, and the leucine-rich activation domain (encoding the nuclear export signal) with only one nonsynonymous mutation present in each (notwithstanding the previously mentioned frameshift in RO7). Again no specific amino acid motif serves to completely distinguish virus from different anatomical sites. It therefore appears that if a strong selection pressure acting upon one of these alternative reading frames has led to alterations in the *env* coding sequence it will not have done so in all the sequences from a single tissue sample.

For further assessment of the relationships between *env* sequences from different anatomical sites a N-J, J-C distance phylogenetic tree was constructed using the *env* (gp160) sequences (Fig 4.9). The most obvious segregation was between brain and lymph node, with two clearly distinct lineages present. Within the brain lineage left temporal sequences were generally separate from those present in the right occipital lobe, with the exception of LT1 and LT2, which grouped with RO sequences. A number of bifurcations were supported by large bootstrap values (>75%) providing confidence in the final tree topology.

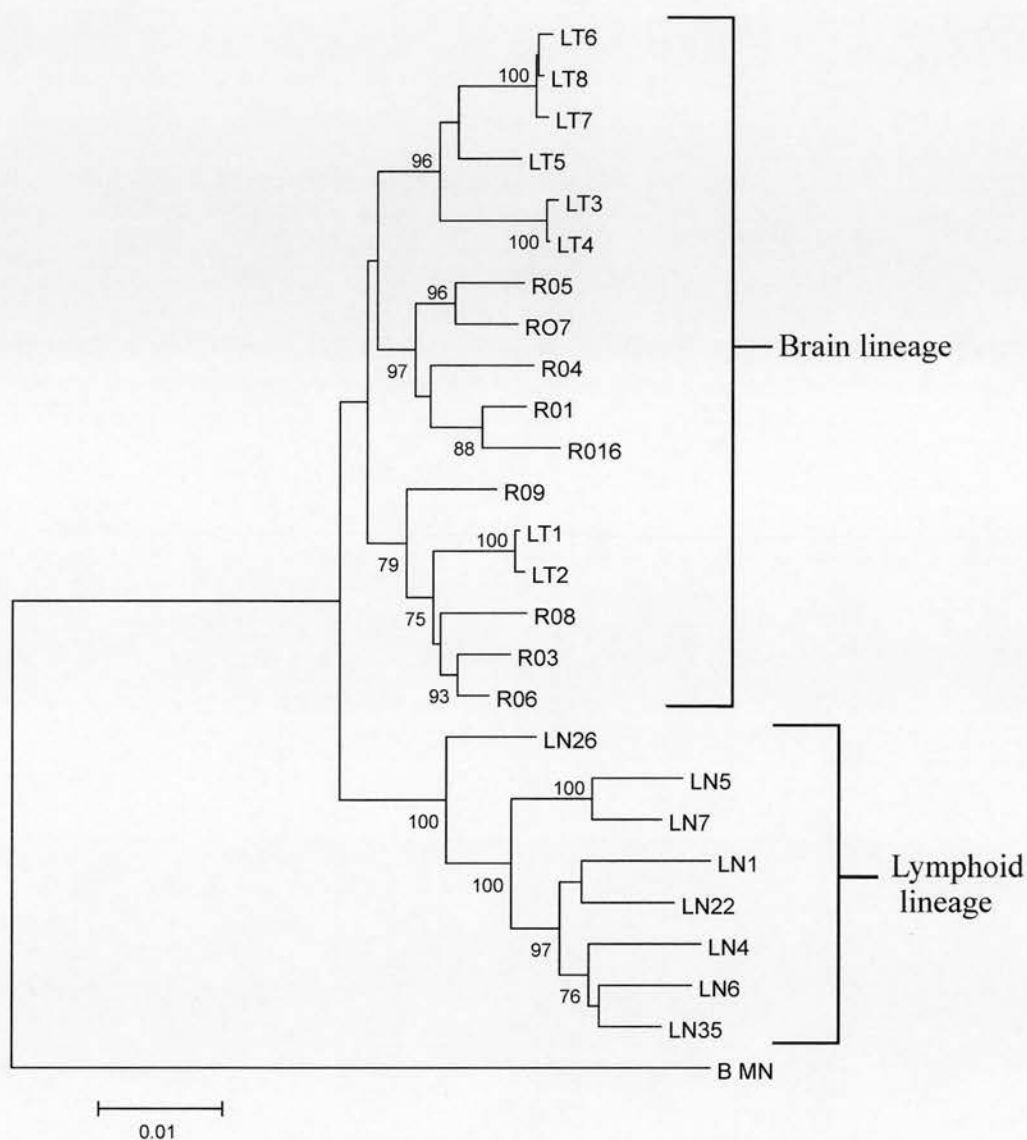


Figure 4.9. Phylogenetic tree constructed with limiting dilution *env* (gp160) nucleotide sequences. The tree was generated as described in the legend of figure 4.5 with the HIV_{MN} sequence used as an outgroup. Bootstrap values over 75% are indicated on branches. RO: right occipital lobe, LT: left temporal lobe, and LN: lymph node. Brain and lymph node sequences group separately and these two major lineages are labelled.

This observation of HIV sequences in the brain being broadly genetically distinct compared to the lymphoid system concurs with preliminary analysis using V3 consensus sequences obtained from the same individual (Chapter 3) in addition to a large number of previous reports. However, such general interpretations provide little information on how the virus present in the brain has acquired the observed variation and whether this apparent segregation is consistent along the length of the sequence as would be expected if representatives of an isolated replicating population had been sampled. An alternative hypothesis as described in Chapter 3 is that a significant amount of recombination occurs during the evolution of regional viral populations. In which case overall segregation between *env* sequences from different anatomical regions may obscure more subtle associations between sequences from different tissues. Closer examination would in that case reveal a more complex interrelationship between variants from different anatomical sites depending upon the specific portion of sequence used for analysis as was found between p17^{gag} and V3 in this and three other infected individuals (Chapter 3).

The relationships between these sequences from different tissue samples were analysed using a new method named "group scanning". This method involves the use of a sliding window moving in steps across the alignment and at each step a grouping value is determined which indicates the amount of association a test sequence has with other user-defined groups in the alignment. The principle is similar to bootscanning (Salminen et al, 1995) although the trees generated are unrooted, so difficulties associated with rooting a phylogenetic tree correctly with an appropriate outgroup acting as the ancestral sequence are removed. Also a bootstrap percentage value is not plotted, and instead for a given placement of a test sequence within the

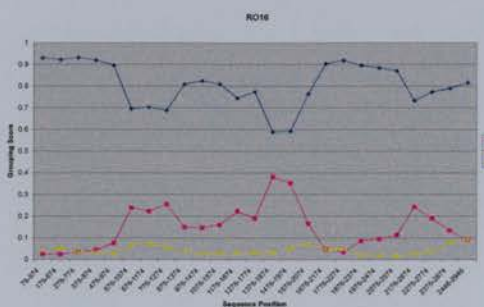
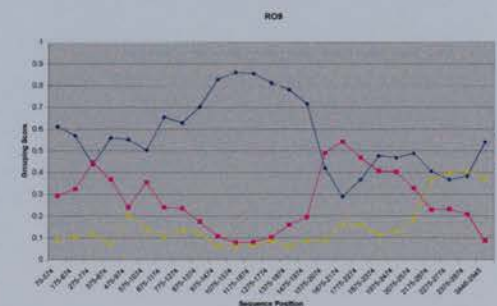
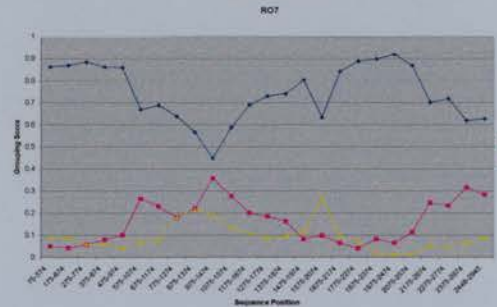
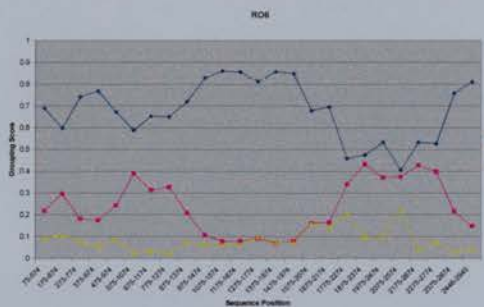
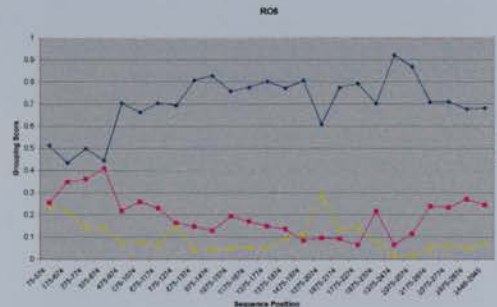
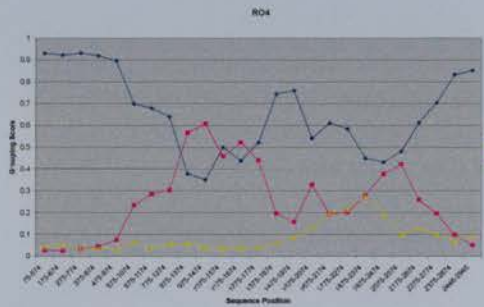
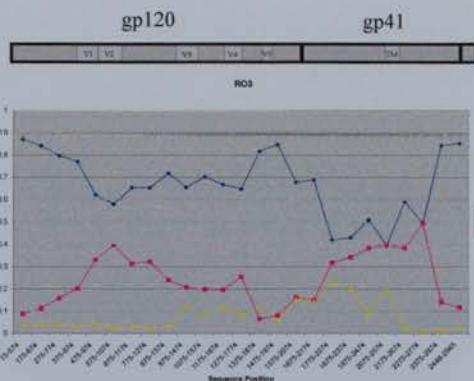
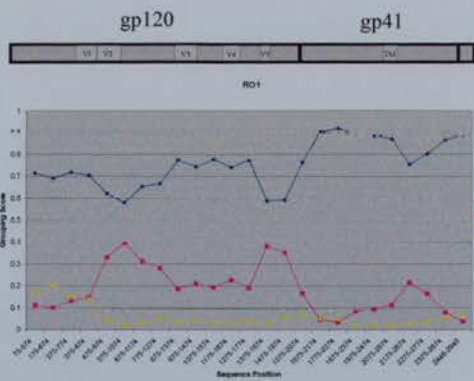
tree a score is assigned to each user-defined group depending on the number of nodes separating this sequence from each other sequence in the data set. Thus the sequence or set of sequences separated from the test sequence by one node is assigned a score of 0.5, the sequence (s) separated by 2 nodes, 0.25, by 3 nodes, 0.125 etc. The sum of scores for each of the individual sequences within a group is calculated. Each group is therefore assigned a part of the total score of 1 based upon its degree of grouping with the test sequence. The tree is then recreated with bootstrap resampling and the process repeated with the average grouping value for each set of bootstrap resampled trees providing the final grouping score of the test sequence with each of the groups of sequences in the data set. The window then slides 1 step and the whole process is repeated across the sequence alignment.

Here the groups were assigned according to tissue (right occipital, left temporal, and lymph node), and a 500 bp window with 50 bootstrap replicates per window and 100 bp steps was used. Each sequence was compared in succession with the remainder of sequences in its own group and all sequences in the other two groups. The grouping score of each individual sequence was then plotted against genome position. The group-scan results for each of the individual *env* sequences are shown (Figs. 4.10 to 4.12).

Sequences RO1, RO3, RO5, RO6, RO7, and RO16 each grouped with other right occipital sequences across the entire *env* alignment, and were more closely related to left temporal sequences than lymph node sequences in the majority of positions (Fig. 4.10).

Figure 4.10 Group-scan plots for right occipital limiting dilution *env* sequences.

Each group-scan plot was generated with a 500 bp window moving in 100 bp steps, with 50 bootstrap replicates in each step (see text for more details of method). Vertical axis represents grouping score, horizontal axis represents window position in *env* sequence. A diagram of Env showing the positions of hypervariable regions within gp120 (V1-V5) and the transmembrane domain of gp41 (TM) is presented above each column.



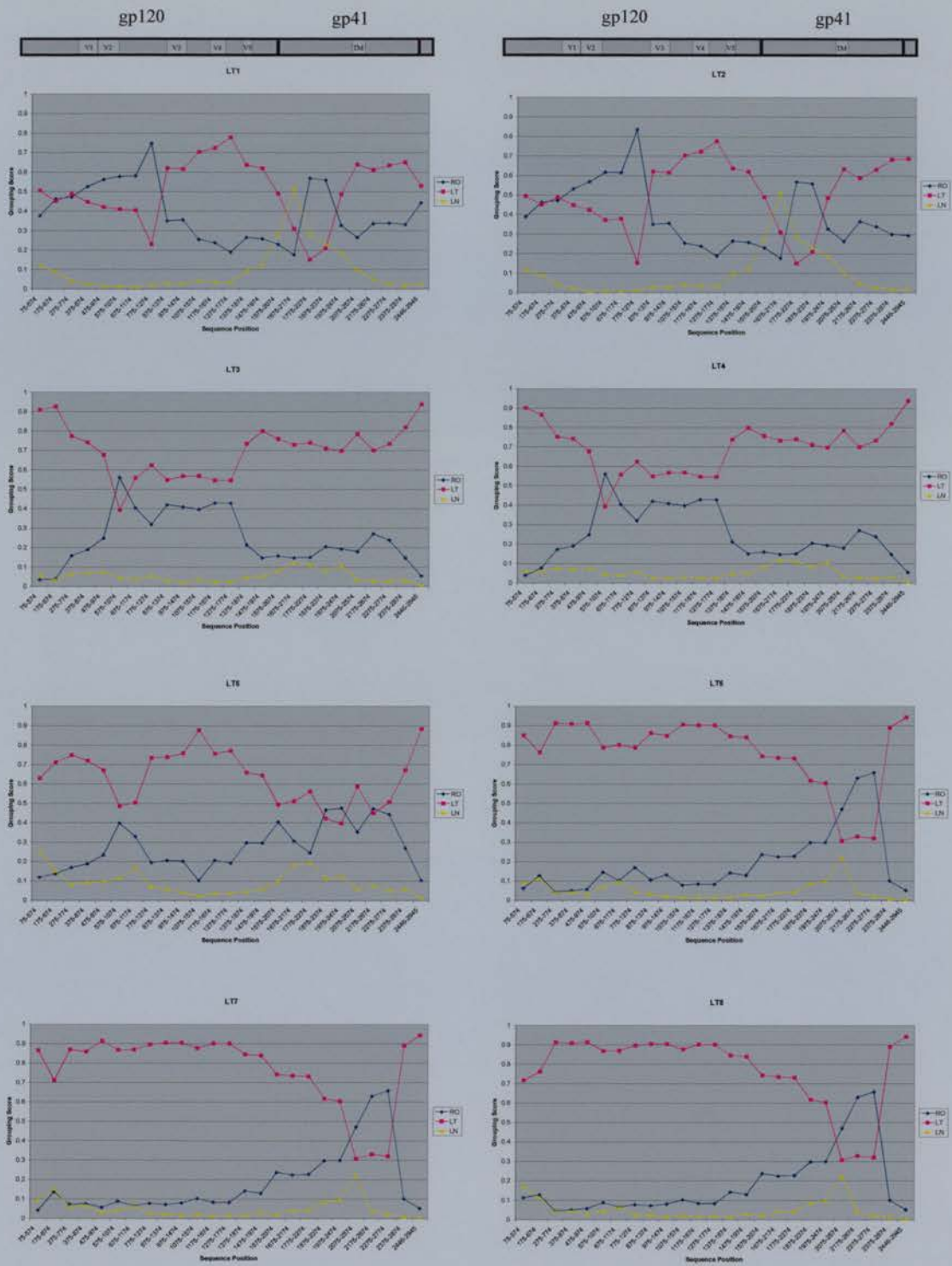


Figure 4.11. Group-scan plots of left temporal (LT) limiting dilution sequences. Method as for figure 4.10.

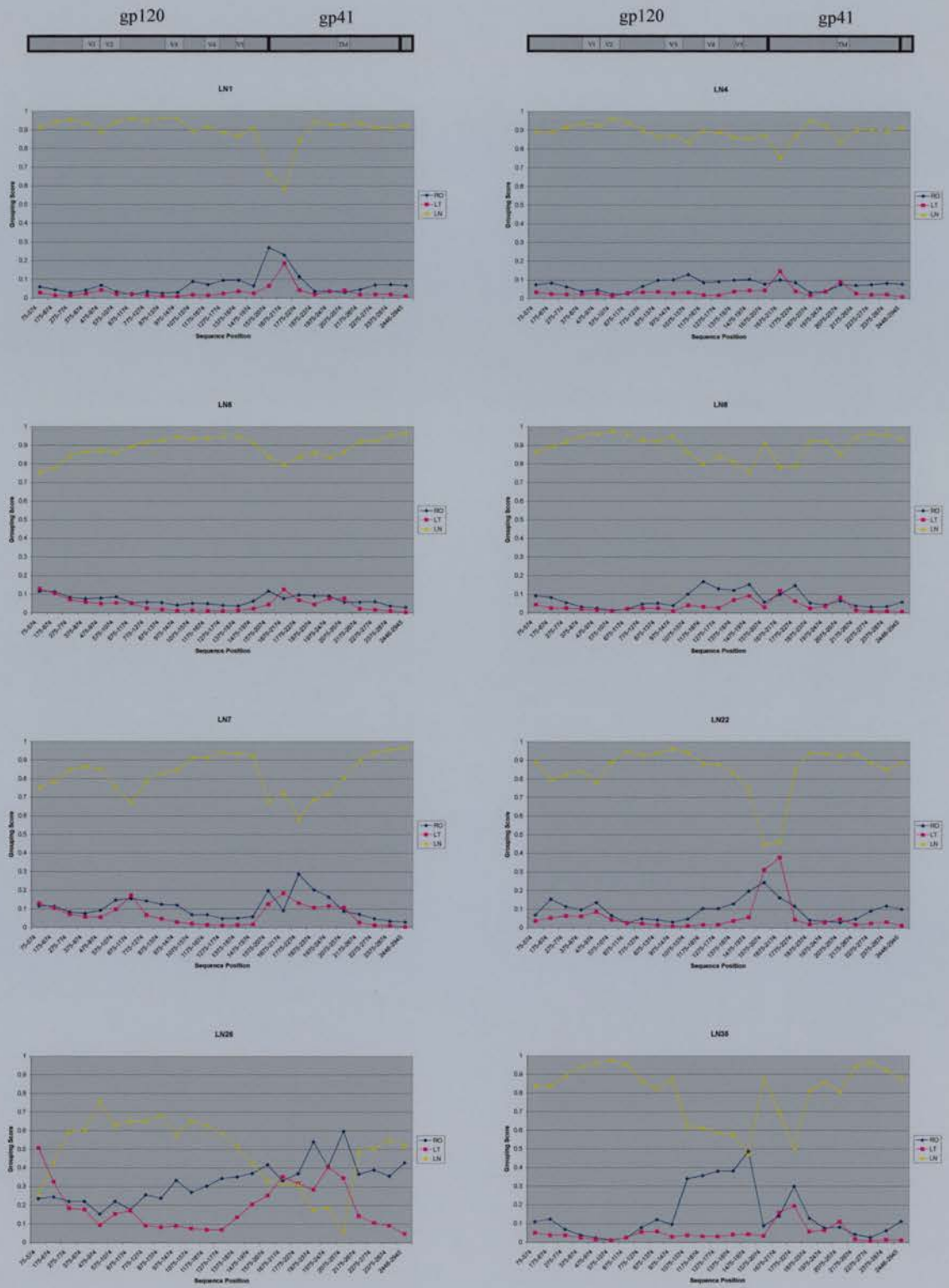


Figure 4.12. Group scan plots of lymph node (LN) limiting dilution sequences. Methods as for figure 4.10

Sequence RO4 was more closely related to left temporal sequences at several points around the V3 region but otherwise grouped with those from right occipital. The most complex interrelationship of a right occipital sequence with those from different tissues was displayed by RO8. This sequence was similar to others from the right occipital lobe until the end of V3, after which it grouped strongly with lymph node until the beginning of gp41. Across the external domain of gp41 RO8 then grouped with either left temporal or right occipital sequences until approximately the transmembrane region where again it groups strongly with other right occipital sequences. RO9 was generally similar to other right occipital sequences apart from at the beginning of gp41 where it grouped with left temporal sequences, and at the end of the Env open reading frame where a brief grouping with lymph node sequences was evident. Left temporal sequences displayed a general association with others from the same tissue and closer relationships to right occipital than lymph node sequences across the majority of the *env* gene (Fig. 4.11). LT1 and LT2 were each more similar to RO sequences from the beginning of V1/V2 to the beginning of V3, then similar to others from left temporal until just after the end of gp120. A brief association with lymph node then right occipital sequences was evident around this point before a return to LT grouping at the transmembrane region of gp41 and for the remainder of the alignment. LT3 and LT4 were similar to other left temporal sequences at all points apart from a single window position beginning in V2 where an association with right occipital sequences was evident. Sequence LT5 grouped slightly more with RO sequences than LT at two positions in gp41, and sequences LT6, LT7, and LT8 each grouped with right occipital sequences in a section of the coding region for the cytoplasmic domain of gp41. Lymph node derived *env*

sequences grouped closely with other LN sequences in almost every case, with low RO and LT grouping scores across the alignment for the majority of LN test sequence comparisons (Fig. 4.12). In marked contrast to other LN sequences, LN26 presented a much more complex set of grouping scores. This sequence associated strongly with LT sequences at the extreme 5' end of the alignment and grouped with both sets of brain sequences more than LN from around V5 to the transmembrane spanning region of gp41.

Several of these sequences therefore appeared to display distinct evolutionary histories depending upon the particular segment of the *env* gene analysed. The most likely explanation for these complex interrelationships is the occurrence of recombination between variants present in different tissues. Inclusion of recombinant sequences in phylogenetic analysis can lead to difficulties because most models for reconstructing evolutionary relationships implicitly assume that recombination has not taken place. The potential recombinants in this nucleotide alignment may therefore have altered the branching pattern of trees where other sequences were being tested. To investigate this possibility, sequences that showed clear inter-tissue association differences across the alignment were excluded from groups, then each sequence was re-analysed against the new groups. No significant differences were observed in the re-analysed group-scans indicating that the possible presence of recombinants had not affected the original results (data not shown). Another potential difficulty in this analysis is that the algorithm for scoring grouping values is new. Therefore necessarily the robustness of this method and has not been determined as thoroughly as older, more established means of determining mosaicism in sequences such as bootscanning (although methods such as those have

their own limitations as described above). To assure that these results were not due to some factor peculiar to this particular method of analysis, a number of sequences were also tested against the same groups using the bootscanning method (see Chapter 2 for details). Results of bootscan analysis were broadly similar to the group-scan results, although some variation reflecting differences in the underlying algorithms and methods was apparent (data not shown). However, these differences were not sufficient to undermine any of the interpretations of group-scan results described here. A representative comparison of group-scan and bootscan results for one apparent non-recombinant sequence (LN4), and one potentially recombinant sequence (LN26) is shown (Fig 4.13). Both methods produced similar (although not identical) plots when used to analyse the same sequence, indicating that the discordant phylogenetic relationships across the *env* gene identified here by group-scanning are valid and representative of those obtained using a distinct method for identification of recombinants. Additionally the amino acid sequence signatures (Fig. 4.7) and synonymous substitutions (data not shown) in the potentially recombinant regions support this interpretation.

The most complex and apparently recombinant sequence relationships were found in RO8, LT1, LT2, and LN26, all of which grouped with sequences from each of the three tissues at some point in the alignment. Interestingly sequence LN26 was present as an outlier group from the main lymphoid lineage in the tree generated using entire *env* sequences (Fig. 4.9).

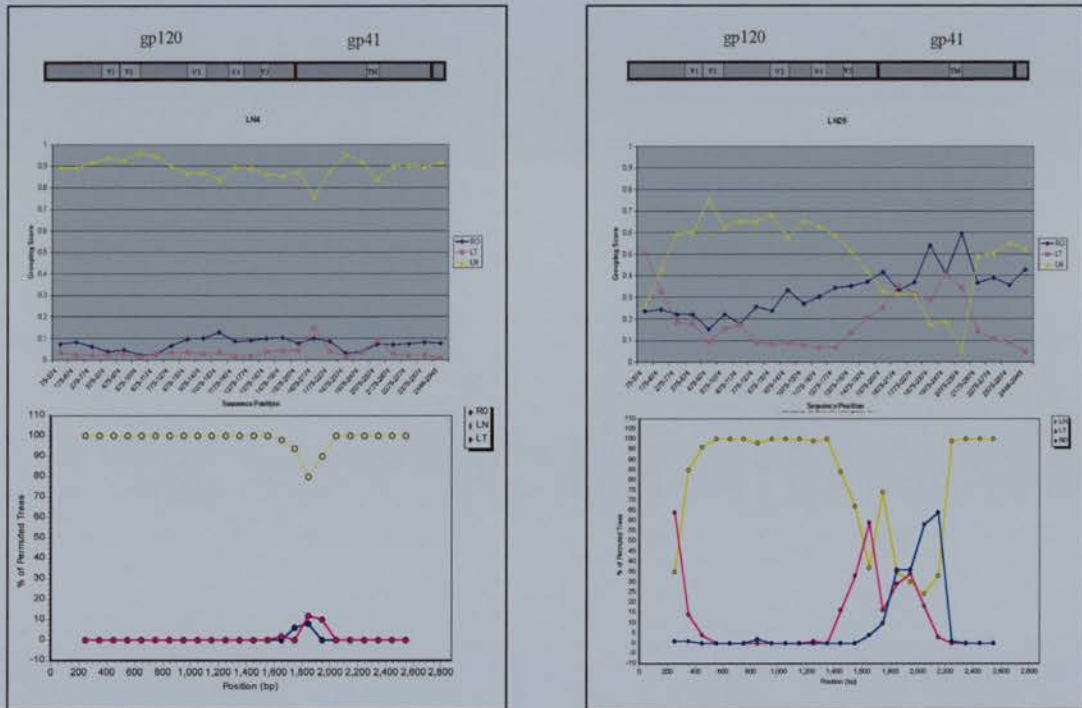


Figure 4.13. Comparison of group-scan with bootscan. Group-scan graphs are shown in grey, bootscans in white. Two sequences analysed with both methods are shown. The left panel shows a sequence that appeared non-recombinant by groupscanning (LN4). The right panel shows a potential recombinant as identified by group-scanning (LN26). In both cases bootscan analysis produced a similar, but not identical plot to group-scan.

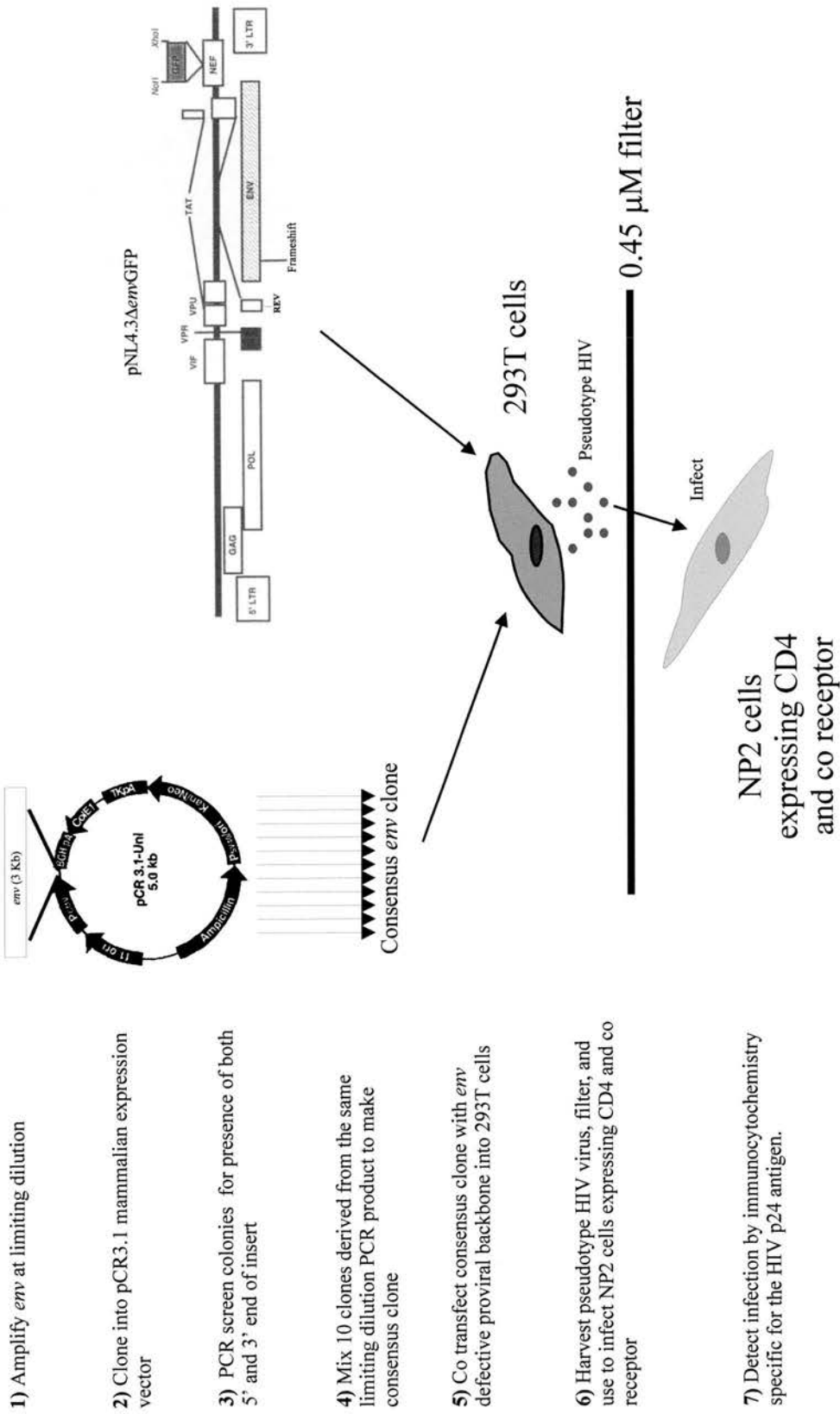


Figure 4.14. Schematic representation of consensus clone and pseudotype virus production.

This type of grouping is characteristic of recombinant sequences, where they do not fit well within defined clades and their final position in the tree is often defined by the proportion of each parental sequence that they contain rather than a clear representation of their evolutionary history. Similarly LT1 and LT2 were the only left temporal sequences to group with those from right occipital in the *env* tree and closer investigation of these two sequences revealed that they contained large sections of right occipital-like sequence. The remainder of the brain sequences grouped only with others from the same tissue, or showed minor grouping swaps between the two brain regions. This minor swapping to an alternative group, where the scores for the two tissues are very similar probably reflects a natural divergence of the test sequence from both groups of brain sequences (i.e. in both RO and LT) at these points, or possibly recombination with a separate lineage not represented in this data set, and does not necessarily indicate a recombination event between members of the three sampled virus groups (RO, LT, and LN). One possible exception is near the 3' end of LT6, LT7, and LT8, where grouping scores between RO and LT sequences are more distinct, thus it is possible that these sequences were generated by recombination. Apart from the identification of recombinant sequences these results demonstrated that intra-tissue association of most LN sequences across the *env* gene was stronger than that within either of the two brain-derived tissue groups. Moreover in most cases non-recombinant brain sequences were more associated with those from the other region of brain than with lymph node sequences.

4.4.6 Consensus pseudotype virus production

In order to functionally characterise envelope sequences, limiting dilution PCR reactions were cloned into a eukaryotic expression vector (pCR3.1) and colonies

were PCR screened to verify that they possessed both the 5' and the 3' ends of the insert and therefore contain the entire HIV *env* open reading frame (ORF). This was necessary because without the screening process in some cases >80% of the resultant colonies contained plasmids with truncated inserts (data not shown), possibly a result of the unintended expression of the *env* sequences in *E. coli* (Cunningham et al., 1993) coupled with the toxicity of this gene product (Cunningham et al, 1993; Peden, 1992) leading to the death or poor growth of cells containing full-length *env* ORFs. Having obtained full-length, recombination free *env* clones, we next devised a strategy to avoid the harmful effect of nucleotide misincorporation on the biological properties of the clones. Although misincorporation cannot be prevented, we prepared consensus clones for biological characterisation that were each composed of 10 individual clones derived from the PCR product of a single molecule amplification (Fig. 4.14). As described above it was empirically determined that individual *env* clones contain a mean of 1.57 misincorporations in the *env* coding sequence. The probability of every one of the ten individual clones combined containing a misincorporation error was therefore reduced to approximately 10.7%. This figure is obtained by applying the average of 1.57 mutations/clone to the poisson formula so: $1.57 = -\ln(f_0)$, where f_0 is the observed frequency of clones with no mutations. Therefore $-1.57 = \ln(f_0)$, and $e^{-1.57} = 0.2$, giving a frequency of clones with no mutations as 0.2 or 20%. The probability of a single clone containing at least 1 polymerase misincorporation mutation is then 80% ($1 - 0.2 = 0.8$), and the probability of every one of 10 clones having an artefactual mutation is 10.7% ($0.8^{10} = 0.107$). In addition the possibility of separate clones containing identical misincorporations is small. In the worst case of a copying error occurring in the first

round of the PCR, only 25% of resulting clones will contain this sequence change, while the possibility of two identical but independent nucleotide mutation events occurring at the same site is remote. Consequently ten individual bacterial colonies, all derived from a single limiting dilution reaction were picked, grown overnight separately to equivalent concentrations, and then mixed in equal proportions before extraction of the plasmid DNA. The validity of this approach in generating *env* gene clones for use in subsequent phenotypic analysis was assessed by the production of pseudotype virus. Consensus clones were calcium phosphate co-transfected into 293-T cells along with NL4.3 Δenv GFP (an *env* defective HIV-1 proviral backbone clone in the vector pSVIII) (Fig. 4.14). Pseudotype virus was harvested after two days and used to infect NP-2 cells (Soda et al, 1999) expressing CD4 and one of the HIV co-receptors CCR5 or CXCR4 in a single round infection assay. Each of the consensus *env* clones produced viable pseudotype viruses that used CCR5 and not CXCR4 in conjunction with CD4 for entry into cells (Table 4.3). However, the virus produced by the consensus clones had approximately 10-fold reduced infectivity when compared to that generated with the control *envs* YU2 and HXB2 (Table 4.3).

Clone*	Cell type†	Average FFU/ml‡
pNL4.3ΔenvGFP only	CCR5	0
	CXCR4	0
YU2	CCR5	9.5×10^4
	CXCR4	0
HXB2	CCR5	0
	CXCR4	3.25×10^4
RO1	CCR5	2.1×10^3
	CXCR4	0
RO3	CCR5	1.75×10^3
	CXCR4	0
RO6	CCR5	4.1×10^2
	CXCR4	0
RO8	CCR5	3.3×10^3
	CXCR4	0
RO9	CCR5	1.27×10^3
	CXCR4	0
RO16	CCR5	4.3×10^2
	CXCR4	0

Table 4.3. Result of pseudotype single round infection assay using consensus *env* clones. *Clones used in conjunction with pNL4.3ΔenvGFP to create pseudotype virus. YU2 and HXB2 were controls for CCR5 and CXCR4 usage respectively. RO refers to consensus clones derived from limiting dilution *env* PCR products. †NP2 cells expressing both CD4 and either CCR5 or CXCR4 were used to determine pseudotype virus co receptor usage. ‡Focal forming units/ml present in the initial harvested media was calculated by taking the average number of foci of infection in the last two wells (in a 10-fold dilution series) showing evidence of infection and multiplying by the dilution factor.

4.4.7 Plasmid rearrangements during expansion in *E. coli*

Detection of large numbers of truncated *env* inserts after initial cloning of PCR products indicated a severe instability of these clones when propagated in *E. coli*. This effect was observed despite the use of different growth media, culture volumes, or growth temperatures for bacterial expansion (data not shown). Three different *recA1* (-) *E. coli* strains were also tested as hosts yet none improved the genetic stability of *env* containing plasmids (data not shown). This occurrence was clearly evident when individual glycerol stocks of each of the ten component clones present in a single consensus clone were prepared, then re-grown from a single bacterium in accordance with good microbiological practice. Original consensus clones (5' and 3' PCR screened at colony stage) used to create the glycerol stocks appeared to be of correct length upon restriction digestion (Fig. 4.15.A) and homogeneous when sequenced. However, after mixing, plasmid extraction, and restriction digestion of clones re-grown from glycerol stocks a range of truncated inserts was present (Fig. 4.15.B). This difference probably reflects the increased number of bacterial and plasmid divisions during culture from a single bacterium without screening as compared with from a colony in which the presence of plasmid containing full-length insert has already been verified. A greater number of cell/plasmid divisions provides additional opportunities for deletion events to occur within the *env* inserts and subsequent selection of the more "fit" bacteria containing these altered sequences.

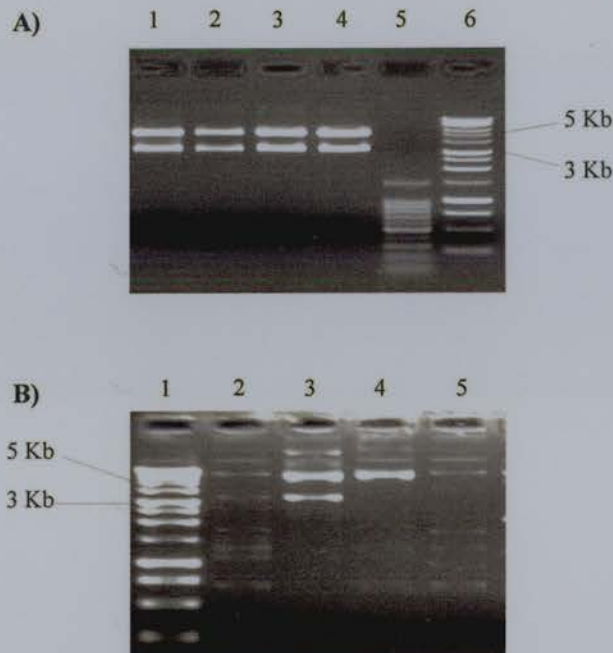


Figure 4.15 Example of *env* clone instability in *E. coli*.

A) *Pme* I digested mini-preps of consensus clones. These clones were PCR screened at the colony stage to verify that they contained both the 5' and 3' end of inserts. The 5 Kb vector and 3 Kb insert restriction fragments both form discrete bands. Lanes 5 and 6 are 100 bp DNA ladder and 1 Kb DNA ladder respectively. Lane 1: RO1, Lane 2: RO3, Lane 3: RO4, and Lane 4: RO5 consensus clones.

B) *Pme* I digested mini-preps of the same consensus clones shown above but after re-growth of a single bacterium from glycerol storage at -70°C . Digests of consensus clones after re-growth display a number of restriction fragments larger than 5 Kb. In some cases this could indicate the loss of one restriction site. Multimerisation of uncut plasmid due to the loss of both restriction sites may explain the presence of visible bands over 8 Kb. A large number of fragments smaller than 3 Kb were also present, probably reflecting the truncation of insert or vector. Lane 1: 1 Kb DNA ladder, Lane 2: RO1, Lane 3: RO3, Lane 4: RO4, and Lane 5: RO5.

The procedure of mixing ten individual clones, grown in separate culture media and derived from a single limiting dilution PCR product is likely to make the instability of these clones more evident, but highlights the complications relating to *env* sequence amplification in *E coli*.

4.5 Discussion

The genetic and functional analysis of primary HIV-1 envelope glycoproteins is central to our understanding of many aspects of HIV biology. This study investigated some of the important artefacts associated with PCR amplification and cloning of HIV-1 *env* genes and provides means of circumventing them. One such issue is the incorporation of errors during amplification by the DNA polymerase used. Single nucleotide misincorporations during PCR amplification of viral sequences has often been mistaken for genuine *in vivo* genetic diversity. Previous investigations into this effect (Smith et al, 1997; Bracho et al, 1998) have identified a number of studies where the majority of nucleotide diversity observed in cloned PCR products amplified from hepatitis C virus (HCV) or HIV RNA sequences falls within that expected from polymerase error alone (Martell et al., 1994; Martell et al., 1992; Najera et al., 1995). Indeed these misincorporations are likely to be present to some extent in the vast majority of studies based upon PCR and cloning. Difficulties arise when counter intuitive results likely to be due to this artefact are presented as genuine sequences. For example as previously identified (Bracho et al, 1998) one study into hepatitis G virus (HGV/GBV-C) phylogenetic relationships reported that the genetic diversity of sequences within a single individual obtained by PCR and cloning was greater than that found by direct sequencing of PCR products from

different patients from around the world. This type of result is likely to reflect PCR misincorporation errors in clones rather than genuine viral genetic diversity and if observed should be verified by direct sequencing of (preferably limiting dilution) PCR products. Clearly if PCR derived clones are used for expression experiments then these difficulties can be transferred to subsequent phenotypic assays, where small changes may lead to significant differences in biological function. The availability of thermostable DNA polymerases with higher fidelities than *Taq* such as *Vent* (New England Biolabs), *Tbr* (NBL Gene Sciences), *Pwo* (Boehringer Mannheim), *Pfu* (Promega), and polymerase mixtures such as Expand High Fidelity PCR system (Boehringer Mannheim) allow DNA amplification with several times the fidelity of *Taq* alone. However, with a large number of thermal cycles often used in nested PCR procedures, and increasing sequence lengths amplified with this method, polymerase misincorporation is still a significant source of artefactual results. In this study the nested PCR amplification of an approximately 3 Kb region including the entire *env* open reading frame using 30 full length HIV-1 genome clones as template resulted in an error rate for Expand High Fidelity Polymerase of 1.51×10^{-5} , or one error in 6.6×10^4 incorporations, compared with the previously reported one error in 1.18×10^5 incorporations (Barnes, 1994). This discrepancy may be due to differences in the specific PCR conditions used in the two studies, or perhaps because of additional alterations incorporated into the *env* sequence during expansion in bacteria (see below). In any case the observed error rate is less than that obtained with standard *Taq* alone (2.6×10^{-5} or 1 error per 3.8×10^4 incorporations). Approximately 64% (7/11) of misincorporations identified in clones from *env* PCR reactions resulted in changes in the predicted amino acid sequence (Fig. 4.1), and

some of these changes may alter the biological function of resulting Env glycoproteins. Potentially important changes included one in the V2 region, and the loss of a cysteine residue essential for maintenance of the V3 loop. Both the V1/V2 and V3 regions of *env* are known to influence cellular tropism and receptor binding of HIV-1, therefore the inclusion of such altered sequences in phenotypic analysis may result in observed properties that are not indicative of the original templates.

Another important issue associated with the use of PCR to amplify genetically distinct templates is that of artefactual recombination. Recombination during PCR has proven to be present wherever heterogeneous sequences are amplified in a single reaction, whether this is in multigene families, repetitive sequences, or heterogeneous RNA virus populations. The effect was first noted soon after the development of PCR (Saiki et al, 1988; Olsen and Eckstein, 1989) and has since been more comprehensively characterised (Meyerhans et al, 1990; Yang et al, 1996; Zylstra et al, 1998; Judo et al, 1998; Fang et al, 1998; Odelberg et al., 1995; Wang and Wang, 1997; Sinkora et al., 2000; Shafikhani, 2002). Observed frequencies of *in vitro* recombination have so far been shown to range from 1% (Judo et al, 1998) to 45% (Sinkora et al, 2000) depending on the specifics of amplification and detection methods. The most important factors appear to be the length of the amplified product, the number of amplification cycles, the DNA polymerase used, and the amount of template DNA initially present.

In the current study mixing either 3000, 300, or 30 copies each of two genetically distinct full-length HIV-1 genome clones followed by *env* PCR and restriction mapping indicated a relatively high frequency of recombination over a 1285 bp

region (Fig. 4.3.A). These results are similar to those of previous studies (Yang et al, 1996; Sinkora et al, 2000) and highlight the significance of *in vitro* recombination when amplifying sequences from heterogeneous virus populations (such as HIV) by PCR. An analogous experiment was performed at near limiting dilution of the HIV_{CAMI}/HIV_{LAI} mix with an average of 2.2 initial template molecules present at the beginning of each reaction (Fig. 4.3.B). In several of the replicate reactions that contained at least one template copy of each genotype this resulted in a restriction pattern indicating the presence of approximately 50% recombinants (the maximum possible with this detection system) implying a complete lack of genetic linkage between the two markers. The same characterisation procedure was performed on replicate *env* PCR products derived from a further dilution of the genotypic mix and resulted in recombinants in 2/9 replicates and the retrieval of single genotypes in 5/9 replicates (Fig. 4.3.C). Thus this study represents the first demonstration that recombination in PCR can occur even when very few template copies are present at the beginning of a reaction. These observations discount the theoretical possibility that *in vitro* recombination is significantly increased by high initial target copy numbers leading to competition of partially elongated transcripts with primers at earlier stages in the reaction (Meyerhans et al, 1990).

One study investigating the importance of retroviral recombination in the acquisition of linked mutations leading to increased viral resistance to zidovudine included information on the background level of recombination due to PCR and cloning procedures (Kellam and Larder, 1995). The authors reported that a 16% artefactual recombination frequency with 100 ng of cloned template DNA became only 5.5% when the amount of input DNA was reduced to 0.1 ng. A subsequent report

describing the amplification of reverse transcriptase for antiviral resistance profiling used this information as a basis to disregard any concerns of artefactual recombination in their PCR procedure (Shi and Mellors, 1997), despite the use of different primers and PCR conditions. The rationale being that if 100 pg of input DNA led to 5.5% recombination in PCR then using 1 pg as in their assay would result in a frequency of less than 5.5%. However, as described here a high frequency of recombination may occur even when only a few amplifiable template molecules are present in the PCR reaction. Thus the frequency of artefactual recombination cannot be predicted simply on the basis of input copy numbers and must be empirically determined for each set of reaction conditions. It is also noteworthy that the recombination frequency documented here using the template clones HIV_{CAM-1} and HIV_{LAI} would be an expected underestimation of that observed when amplifying many different but more closely related DNA templates such as may be present in a single sample from an HIV infected individual. Recombination during PCR is especially significant in the context of the HIV *env* gene because HIV is by nature highly heterogeneous and undergoes frequent recombination *in vivo* (Morris et al, 1999; Hung et al, 1999). Therefore *in vitro* recombination has traditionally been difficult to differentiate from natural variation and evolution *in vivo*. Undoubtedly the type of sequence scrambling observed here occurred to some extent during each of the previous attempts to characterise *env* sequences amplified by PCR and may have strongly affected both genetic and phenotypic analysis. The PCR amplification of genetically distinct template molecules has been suggested as an alternative to the more technically demanding "DNA shuffling" (Stemmer, 1994a; Stemmer, 1994b) approach to recombination mutagenesis (Judo et al, 1998). This approach is used

during *in vitro* evolution experiments and involves the generation of chimeric molecules for subsequent selection of those with altered biological properties from the parental phenotypes. It may be considered somewhat paradoxical that the same procedure utilised in some areas of biological science for generating novel chimeric molecules is also routinely used in HIV virology for amplification of sequences that are intended to accurately reflect virus *in vivo*. Methods have been outlined for the reduction of *in vitro* recombination (Zylstra et al, 1998; Judo et al, 1998; Fang et al, 1998) yet these invariably advocate very low numbers of amplification cycles, a practice that is likely to result in a significant reduction in PCR sensitivity making the procedure unsuitable for amplifying viral sequences from all but bulk cell or tissue DNA extractions. More importantly there has been no study describing the complete elimination of this artefact by any method other than limiting dilution.

Restriction analysis of limiting dilution *env* PCR products demonstrated that unadulterated single genotypes could be retrieved from an initial genotypic mix and therefore the potential for *env* amplification from single proviral copies. This was confirmed by comparing the *env* PCR sensitivity with that of a V3 PCR using both cloned HIV-1 DNA as template (Table 4.1) and proviral DNA isolated from tissue obtained at autopsy from the right occipital lobe of an HIV infected individual with evidence of encephalitis (Table 4.2). These results are consistent with those of a study using similar conditions for HIV-1 *env* PCR, which found that the sensitivity achieved was such that it could be used to amplify the *env* gene from a single provirus (McClure et al., 2000). The direct sequencing of PCR products amplified at limiting dilution provides a method to prevent both single nucleotide

misincorporations and recombination in PCR and is therefore an extremely useful approach for accurate genetic analysis of HIV-1 *env*.

Use of limiting dilution PCR has several advantages in addition to the elimination of *in vitro* recombination and sequencing errors due to nucleotide misincorporation. Firstly, preferential primer binding to certain viral genotypes present in an original mixed population can lead to an over representation of these sequences in the end product of a PCR reaction (Kwok et al., 1990). If only one initial sequence is present in the mix then there is no opportunity for “swamping” of the PCR reaction by other variants. Another common problem with traditional cloning and sequencing techniques is that of re-sampling of clones (Liu et al, 1996), this is the process whereby several product clones from a sample are derived from the same initial provirus. Re-sampling can be due the presence of few proviral copies in the initial specimen, or because of disproportionate amplification of certain templates in the PCR reaction as described above.

To formally address the issue of resampling frequencies, the sequence diversity obtained by direct cloning using an adequate number of template molecules was compared with that of limiting dilution PCR on the same DNA sample. The V1/V2 region of *env* was amplified using over 1000 proviral molecules as template then the resulting PCR product was cloned. Eleven isolated clones derived from this single PCR reaction were then fully sequenced. The probability of a single resampling event occurring with this number of input molecules was less than 5.4%. For comparison 9 limiting dilution *env* PCR products were amplified from the same DNA sample and the V1/V2 region directly sequenced. Upon phylogenetic analysis,

cloned sequences displayed reduced diversity compared to those obtained by limiting dilution and close inspection of the nucleotide sequences revealed that clones contained a median of 1 mutation per sequence compared with 7 for PCR products. The reduced diversity was largely due to at least four of the cloned sequences being representative of only a single original proviral sequence. These results therefore demonstrate that inputting a large number of template molecules into a PCR reaction does not guarantee that clones derived from that reaction would not represent a restricted number of original sequences. One way to partially circumvent this effect is to perform several separate PCR reactions from the same stock of sample DNA, and mix these PCR products prior to cloning thereby assuring that resulting clones reflect at least several original template molecules. This approach is useful when sequences appear extremely similar and resampling of clones is suspected. For example it has been used in a study examining HIV-1 envelope sequence evolution in the setting of potent antiviral therapy (Gunthard et al., 1999), where a large proportion of clones obtained from the C2-V3 region of *env* in three patients undergoing antiretroviral therapy appeared identical. If the clones had not been obtained from a mixture of 4 separately amplified PCR products then resampling error would have been strongly suspected. However, this method of mixing separate PCR products prior to cloning does not prevent other problems such as recombination in PCR and the incorporation of single nucleotide misincorporations.

A recent report (Liu et al, 2000) using a very similar *env* PCR to that described here almost certainly contained cloned *env* sequences that were the result of resampling. The report focussed upon *env* sequences obtained at autopsy from bone marrow, lymph node, lung, and four regions of brain in addition to samples of blood

monocytes collected 5 months prior to death of the study subject. Phylogenetic analysis demonstrated that cloned sequences obtained from the brain deep white matter grouped more closely with bone marrow and monocyte sequences than those in the remaining areas of brain and other tissues. The authors interpretation of this was that HIV infected bone marrow derived monocytes had entered the parenchyma of the deep white matter, where they either transmitted infection to this site or were retained themselves as perivascular macrophages. This evidence was used to propose that increased or initiated monocyte trafficking into the brain occurred during late stage infection and is a critical step in the development of HIV dementia, an idea that was later expanded in a review which cited the same sequence data (Gartner, 2000). While this underlying hypothesis is plausible, there are a number of reasons for believing that the sequences from deep white matter (5 clones), monocytes (4 clones), and the head of the caudate (5 clones) are each only representative of a single provirus present in the original DNA samples. Firstly, no specific efforts were made to avoid resampling of clones and no indication of the proviral titre in each of the samples was given, although as the brain appeared grossly normal upon pathological examination with no multinucleated giant cells present it is probable that the viral load was low (Bell, 1998; Gray et al, 1996). Members of each group of cloned sequences were also extremely similar with a mean number of nucleotide differences/clone of 1.8 for deep white matter, 1.5 for monocytes, and 1.6 for head of caudate. These figures are very similar to the 1.57 mean nucleotide mutations/clone described here after *env* PCR from cloned template DNA. Furthermore all of these identified differences were sporadic mutations not present in other clones, a characteristic of polymerase misincorporation during PCR. Clone F7 from head of

caudate did contain an additional single frameshift mutation which may genuinely separate it from other sequences in the same sample, although such frameshifts can also result from polymerase slippage in PCR (Viguera et al., 2001) most frequently at sites of DNA secondary structure.

In both this and the described study the Expand High Fidelity PCR system (Boehringer Mannheim) was used for amplification of *env* by PCR, therefore a similar number of polymerase misincorporations would be expected in each. Moreover the *env* gene constitutes over a quarter of the viral genome, and HIV-1 undergoes approximately 1 mutation per replication cycle (Roberts et al, 1988; Preston et al, 1988; Ji and Loeb, 1992; Mansky and Temin, 1995). A replicating HIV population would thus be expected to contain at least a few mutations across the entire *env* gene above the background number of polymerase misincorporations that occur during PCR, even if these mutations arose solely by genetic drift. It is possible that expansion of an infected lymphocyte resulted in the reported identical cloned sequences, although this is unlikely to have occurred so uniformly in several separate anatomical sites. The isolation of identical sequences (probably all derived from a single proviral template) in PCR reactions using 1 µg of template DNA extracted from tissue or cells (approximately 1.5×10^5 cells) as described in the study is consistent with contamination of samples by a small number of infected cells from peripheral blood. If this is the case then it is probable that any observed phylogenetic groupings reflect the stochastic nature of sampling such small sequence numbers rather than genuine biological events. This example of resampling error is not unique, and the presence of resampling in sequence data sets can cause erroneous estimations of viral diversity that may result in quite misleading conclusions

regarding the tissue specificity of HIV. Indeed, the typical tree topologies obtained when re-sampled sequences are included in phylogenetic analysis appear superficially as one would expect if a genetically distinct "tropic" population of virus were present within the sample, with large genetic distances between statistically supported clades and very little diversity within them.

Although outwith the scope of the current study it worth noting that the amplification of heterogeneous templates by PCR can also lead to other downstream difficulties. If heteroduplex molecules are present at the end of a reaction and the product is cloned, then *in vivo* repair in *E. coli* may also lead to the formation of chimeric molecules, possibly by a mechanism involving mismatch repair (Shafikhani, 2002). This provides further evidence advocating the use of limiting dilution PCR.

The obvious potential for misinterpretation of sequence data that may have been subject to one or more of the *in vitro* artefacts described above clearly must be taken into account when obtaining sequences for genetic analysis. Therefore limiting dilution PCR was used to amplify *env* sequences for further investigation into the possible tissue specificity of HIV-1. Nine limiting dilution *env* sequences were obtained by this method from the right occipital lobe of brain, 8 from the left temporal lobe, and 8 from a lymph node of subject NA021. V3 sequences from the different regions were relatively similar with only 3 amino acid positions varying. Two variable sites in the *env* alignment were perhaps indicative of adaptation to infect cells in the brain. One of these was at position 301, an important residue for CD4 binding (Kwong et al, 1998). Each of the brain sequences contained an asparagine at this position as opposed to a valine or alanine in lymph node-derived

sequences. Segregating mutations between brain and lymph node-derived HIV sequences such as this (in CD4 or CCR5 binding domains of gp120) may be expected if adaptation to infect cells in the brain expressing lower levels of CD4 and CCR5 had occurred. Another possibly significant site where all but three of the brain-derived sequences were distinct from lymphoid variants was position 215 in the V1/V2 stem. Brain isolates had lost an N-linked glycosylation site in this position (in a relative asparagine to aspartic acid change). The loss of this site had only been documented in 7 of 208 subtype B *env* sequences in the Los Alamos National Laboratory HIV Database and was therefore highly unusual. Loss of the same glycosylation site was found to alter the position of the V1/V2 variable loops in HIV-1 ADA Env, exposing the CCR5 binding site in gp120 and facilitating CD4-independent infection (Kolchinsky et al, 1999; Kolchinsky et al., 2001b). Gorry and colleagues also used brain tissue from subject NA021, the same individual described here (Gorry et al, 2002a; Gorry et al, 2001) (designated UK1-br in their studies), for amplification of *env* sequences that were subsequently cloned and used in functional analysis. The lacking N-linked glycosylation site at position 215 (corresponding to position 197 in HXB2) described here was also identified in the 3 clones from the brain of NA021 in their study. To investigate whether it contributed to increased CCR5 affinity and/or CD4-independent CCR5 binding they restored this asparagine in one of their clones by site-directed mutagenesis (Gorry et al, 2002a). However, this restoration had no effect on the level of CCR5 binding in the presence or absence of soluble CD4, suggesting that loss of the glycosylation site at this position led to a CD4-independent phenotype only in a strain specific manner and was not applicable to the UK1br-15 clone from NA021. Loss of a glycosylation site at position 215 may

alone not be sufficient to confer increased CCR5 affinity. Yet the presence of an asparagine at this site in all lymph node variants isolated from the same individual provides some evidence for its potential importance in the context of this infection. Given that the glycosylation site was present in virus elsewhere in the infected individual and appeared to be specifically lost during viral replication in the brain, this mutation is suggestive of specific adaptation. The same report comprehensively characterised the brain-derived Envs and found that they induced both high levels of fusion in monocyte-derived macrophages and neuronal apoptosis. In addition these Envs also bound to CCR5 in the absence of soluble CD4 and required only low levels of CD4 and CCR5 expression to mediate membrane binding and fusion in cell-to-cell fusion and single round infection assays, although they could not mediate CD4-independent infection. Virus in the brain of this infected individual therefore appeared to represent a well-adapted, “neurotropic” HIV population. When coupled with previous results (Chapter 3) implying that recombination had occurred between p17^{gag} and V3 in this virus subpopulation, HIV variants from this study subject appeared suitable for further investigation into the potential role of recombination in the evolution of HIV during viral expansion into new anatomical sites.

Obtaining entire *env* sequences at limiting dilution allows direct comparisons between the evolutionary histories of different subgenomic regions to be made, as opposed to these relationships being indirectly inferred using different small sections of sequence obtained individually from the major viral variants present in each anatomical site. Unfortunately, in most areas of evolutionary biology recombination is notoriously difficult to identify unequivocally. One complication is that the original ancestral strains may no longer be extant, and their contribution to

contemporary organisms may be only small sections of sequence. Another is the problem of defining which organisms represent ancestral genotypes and which are recombinants. Taking the example of HIV-1 group M subtypes, strains grouping with the circulating recombinant form CRF_cpx (CY032) (previously designated subtype I) are complex mixtures of subtypes A, G, K, and other unclassifiable areas. Yet if the subtype I mosaic sequence had been isolated before subtypes A, G, and K then this would have been defined as a subtype and A, G, and K termed complex recombinants because they each contain regions of sequence similar to subtype I. Defining a subtype in terms of its relative abundance is also unsuitable because issues such as founder effects can lead to a recombinant becoming the most common genotype in a population, for example the widespread presence of an A/E recombinant now designated CRF01_AE (CM240) (Robertson et al, 2000) in South East Asia (Gao et al, 1996b). Thus “recombinant” must always be used as a relative term unless the genotype of ancestral sequences is known. These difficulties are compounded in the context of variants that have diverged within a single infection where genetic distances are much smaller than found between subtypes and ancestral sequences may be present in their original form as (probably defective) proviruses. However, it is possible to assess the evolutionary relationships of different sections from a single sequence with others isolated from several anatomical sites. If these relationships differ across the sequence then the occurrence of recombination between variants can be inferred. Methods for identifying intersubtype recombination such as bootscanning (Salminen et al, 1995) were originally designed for the classification of recombinant genomes with larger genetic distances between original variants than generally found within single infected individuals. Issues such

as the specific position of an outgroup become more important when similar sequences are compared, therefore a new method based upon branching patterns of unrooted trees coupled with a sliding window across the alignment was developed for analysis of the *env* sequences isolated in this study.

Several *env* sequences were identified in this analysis where different sequence sections had distinct evolutionary histories (Fig. 4.10 to 4.12), most likely as a result of recombination. Thus providing further evidence for the frequent occurrence of this mutational mechanism during HIV replication in the brain. These recombination events did not necessarily occur directly between variants in the right occipital, left temporal, or lymph node but possibly between HIV sequences closely related to those present in two or more of these separate sites. All sequences were predicted to use CCR5 on the basis of negatively charged or uncharged residues at positions 11 and 25 of the V3 loop, and this was confirmed with several RO Envs by construction of pseudotype virus (see below). More subtle Env determinants of tropism may be both complex and strain dependent, and are incompletely understood. Therefore the question of whether the specific recombination events described here have any functional significance is open to conjecture. None of the recombinant *envs* from brain had lymphoid-like portions of sequence in V1/V2 or V3. It may be speculated that progeny virus of such recombination events would be selected against, because if viral adaptation occurs during infection of the brain then these subgenomic regions are likely to contain at least some of the determinants of such tropism. The presence of two right occipital like segments in sequences LT1 and LT2, including the V1/V2 region, and a left temporal like V3 region in RO4, perhaps reflects the interchangeability of these portions of sequence within brain adapted virus. This,

along with other close associations and potential recombination events between the RO and LT sequences suggests both the trafficking of virus between these two regions of brain and the presence of similar selection pressures on viral Env sequences in both of these sites. Such close relationships between virus present in different brain regions also suggest that multiple re-seeding events from the lymphoid system are not solely responsible for the origin and continued presence of virus in different regions of brain.

A number of *env* sequences obtained from brain (RO8, RO9, LT1, and LT2) contained segments that grouped more with those from lymph node than with other brain sequences. The most obvious explanation for this is that lymphoid HIV variants had entered the brain of this individual at some point and recombined with the HIV population already *in situ*. Possible implications of such an event have been discussed (Chapter 3). However it is worth restating that this provides strong evidence against anatomical segregation between virus in the brain and lymphoid system as an explanation of the frequently observed genetic differences between HIV variants in these two sites. If temporal or spatial separation of variants at different anatomical sites were significant factors in generating the observed genetic segregation then other regions of the *env* gene would not be so easily interchanged as described here. Also, recombination of this type provides a powerful mechanism for HIV in the brain to selectively exchange advantageous characteristics such as antiviral resistance or immune escape mutations with variants present in the peripheral blood while maintaining their neuroadapted phenotype. All but one lymph node sequence grouped strongly with others from lymph node, and this intra group association was much more defined than within either of the brain sequence sets.

One possible explanation is that infiltration of infected cells into the brain is relatively common and there are greater opportunities for (the more abundant) lymphoid virus to affect brain populations than the reverse. LN26 contrasted strongly with the other lymph node-derived sequences by containing sections of brain-like sequence. This sequence may represent a lymphoid variant that entered the brain and recombined with virus already at this site, followed by the return of progeny virus derived from this recombination event into the lymphoid system at a later time. Alternatively recombination between a lymphoid variant with a brain-like and probably macrophage tropic variant elsewhere in the body (such as the colon) may have resulted in the observed genetic reshuffling. If the former is true then it lends support to the possibility of HIV from the brain of re-seeding of the lymphoid system in patients upon cessation of antiviral therapy.

These results are derived from a single patient and while they are consistent with those obtained in a similar but less comprehensive analysis of HIV subgenomic regions in several other infected individuals (chapter 1) (Morris et al, 1999), and a large number of other observations regarding recombination in HIV-1 (Jetzt et al, 2000; Sabino et al., 1994; Robertson et al, 1995; Gao et al, 1996b; Jung et al, 2002) their generality has not been fully assessed. Also, the assumption made here is that these recombination events occurred in the brain. However, none of these results preclude the possibility that apparently brain specific sequences and recombinants were originally generated in another anatomical site or cell subset followed by multiple seedings of the brain by virus this site. Therefore the hypothesis of a genetically distinct HIV population in the bone marrow being responsible for multiple seeding of the brain during late-stage HIV infection, facilitated by increased

entry of blood monocytes into the brain parenchyma (Liu et al, 2000; Gartner, 2000) cannot be discounted on the basis of this data. Yet the incompletely defined specific anatomical sites in which the recombination events documented here occurred does not detract from their potential biological significance in the evolution of viral variants during a single infection. Importantly, had these *env* genes been obtained by standard PCR and cloning procedures rather than by limiting dilution PCR then recombination during PCR would have rendered any meaningful analysis of genuine *in vivo* recombination events impossible. Likewise if these *in vivo* recombination events had brought together new combinations of sequences leading to altered Env phenotypes then these too would probably have been lost during PCR amplification of mixed proviral templates.

In order to circumvent each of the potential artefacts associated with PCR and cloning of *env* genes described above, and to generate unadulterated *env* clones suitable for subsequent phenotypic analysis a novel approach was undertaken. Using standard cloning procedures the majority of *env* product clones are likely to contain polymerase errors, even when a mixture of enzymes resulting in fairly high fidelity is used such as described here. However, if ten clones are taken from the product of a limiting dilution PCR reaction then there is a high probability that some will contain no errors. In addition those errors that are present most likely will have occurred at different nucleotide positions, because even if an error occurred in the first round of a then PCR it will only be present in a quarter of the product clones. Moreover only around 75% of the random nucleotide misincorporations will result in amino acid changes in encoded proteins. Therefore the overall sequence of this consensus clone will be identical to the one that was present *in vivo* and consequently if the consensus

clones are used for translation then a significant proportion of the expressed protein will be phenotypically identical to that present *in vivo*. The validity of this approach was assessed by the generation of pseudotype HIV virus. Co-transfection of several *env* consensus clones with an *env* defective proviral clone (pNL4.3 Δ *env*GFP) into 293T cells produced virus that could be used in a single round infection assay. Pseudotype virus generated in this way was viable and used CCR5 in conjunction with CD4 to enter cells (Table 4.3).

In summary the focus here on generating *env* sequences for analysis that are identical to those present *in vivo* is particularly relevant at the present time where interest in the field has shifted from simply identifying the receptors used by isolates to enter different cell types to more subtle considerations. Identification of different molecules that can be used as receptors by HIV-1 primary isolates (both known and as yet undiscovered) is still an area of research that may yield interesting results. Yet observations that the majority of primary viral isolates use either CCR5 or CXCR4 has led to a renewed interest in other factors that may mediate the perceived differential cellular tropism of HIV. For example by identifying not only the type of receptor used but also the density of and conformation of these molecules required by different primary viral isolates to adsorb and fuse with the host cell membrane. The binding affinity of Envs to CD4 or coreceptors and the specific regions of the coreceptor molecules used by distinct primary virus Envs is also under close scrutiny. This type of analysis when used in conjunction with the careful profiling of HIV-1 receptor expression patterns in different primary cell types, and *in vitro* infection experiments using chimeric or pseudotype viruses, will provide a more thorough understand of the *in vivo* tropism of HIV-1. A crucial consideration when

performing these technically elegant investigations is to ensure that the viral sequences used for analysis are a true representation of the original *in vivo* virus and not the product of *in vitro* artefacts. Several of such artefacts were highlighted and quantified here and means of eliminating them provided. HIV-1 envelope sequences amplified at limiting dilution from the brain and lymphoid tissue were, overall, genetically distinct, implying the presence of an adapted viral population in the brain. The identification of segregation between brain and lymphoid isolates at two positions in the *env* gene with potentially important roles in receptor binding supports this assessment. However, such segregation is not evident across the *env* gene and the evolutionary relationships between variants present in the right occipital lobe, left temporal lobe and lymph node of this individual varied depending upon the specific region of *env* analysed. Such complex interrelationships imply that recombination is a frequent event during the evolution of HIV in the brain and that anatomical separation of variants in the brain and the lymphoid system is far from complete. Therefore genetic differences between HIV in the brain and the lymphoid system are likely to represent adaptive viral evolution.

Finally, a severe instability of clones containing the *env* open reading frame was also identified during the course of this study (Fig. 4.15), both during initial cloning procedures and after the re-growth of clones from glycerol stocks containing previously verified complete *env* inserts. This observation is consistent with several previous studies indicating that lentiviral *env* gene clones are unstable when propagated in *E. coli* (Peden, 1992; Cunningham et al, 1993; Wang and Mullins, 1995), possibly as a result of uninduced expression of the toxic *env* gene product in bacterial cells (Cunningham et al, 1993). Instability may be manifest as point

mutations, insertions, deletions, and sequence rearrangements (Cunningham et al, 1993). It is entirely possible that these issues may lead to the generation or preferential isolation of *env* clones with stop codons or frameshifts. Such changes were perhaps present in previous studies and could have direct implications in for example the estimation of frequencies of defective provirus and their significance in HIV infection. In extreme cases *envs* with strong cryptic *E. coli* promoter activities or particularly toxic gene products may be completely refractory to isolation by standard cloning procedures. Given the focus here on obtaining genuine *in vivo* sequences the unpredictable alteration of *env* sequences in *E. coli* is far from ideal. A more suitable approach in cases like this would be to dispose of all steps involving *E. coli* in the production of open reading frames suitable for expression in eukaryotic cells. This is the subject of Chapter 5.

Chapter 5: Results (iii)

5.1 Introduction

Efficient expression of recombinant proteins in eukaryotic cells requires the use of DNA for transfection that contains an upstream promoter and a downstream polyadenylation (poly [A]) signal flanking the gene of interest. These sequences allow transcription and translation of the open reading frame when large amounts of DNA (typically >1 μg) are transfected into eukaryotic cells. In general, to produce this DNA a gene is first cloned into an expression plasmid then expanded in *E. coli*, yeast, or other host cell before re-purification. The genetic instability in clones containing the HIV-1 *env* gene during propagation in *E. coli* described in Results chapter 4 strongly suggested that this method was not ideal as a means of generating DNA for subsequent phenotypic analysis of primary HIV-1 Envs. This identification of deletions or rearrangements in *E. coli* of plasmids containing specific inserts is far from unique.

One of the difficulties in cloning the cystic fibrosis transmembrane conductance regulator (CFTR) was the presence of a cryptic bacterial promoter within the CFTR gene sequence. The result was unintended expression of the toxic gene product in *E. coli* and exclusive isolation of extensively rearranged clones. A stable plasmid containing the complete gene was only obtained by the construction of a complementary DNA in a low-copy-number vector (Gregory et al., 1990). Similarly the presence of a cryptic bacterial promoter in the hepatitis C virus (HCV) genome led to the preferential cloning of defective viral sequences, with stop codons or prematurely truncated inserts present in the vast majority of isolated clones (Forns et al., 1997). Insertion of extra nucleotides upstream of the start codon of the HCV

insert led to a significant increase in the number of fully functional clones. Nonrandom selection of clones from an initially diverse population such as in this case can lead to false interpretations of virus diversity and an erroneous estimation of the frequency of defective viral genomes *in vivo*. In the context of lentiviruses a number of previous reports have demonstrated that both full-length proviral clones and individual *env* gene clones are unstable in *E. coli*. One report identified a large number of deletions when amplifying HIV proviral clones in high copy number plasmids, and suggested the use of medium-copy-number plasmids to circumvent this difficulty (Peden, 1992). A subsequent report demonstrated that medium-copy-number vectors were also unsuitable to maintain the integrity of the simian immunodeficiency virus (SIV) and equine infectious anemia virus (EIAV) *env* clones and advocated the use of low-copy-number plasmids instead, although this greatly reduces plasmid yield (Cunningham et al, 1993). The same report also identified a cryptic bacterial promoter sequence present at the beginning of lentiviral *env* genes, leading to the unintended expression of these toxic gene products. The toxicity of membrane protein genes including fragments of HIV-1 *env* cloned in *E. coli* has been well documented (Rose and Shafferman, 1981; Samuel et al., 1988). It is therefore unsurprising that if basal levels of lentiviral Envs are unintentionally expressed in *E. coli* then they would have a toxic effect leading to the preferential isolation of altered plasmids or defective inserts. Thus it is entirely possible that cloning and propagation in *E. coli* of HIV-1 *env* genes leads to the preferential isolation or *de novo* generation of coding sequences with stop codons, deletions, frameshifts, and other functionally deleterious mutations. This issue may have been a factor in the large number of apparently non-functional *env* open reading frames identified in previous studies

based upon cloning of full-length HIV-1 *env* genes (Gorry et al, 2001; Liu et al, 2000).

In addition to the unintended expression of toxic gene products, a number of other factors may make initial cloning procedures difficult and time consuming. For example unusual G+C content or the presence of internal repeats (Yamada et al., 1995) may also make the introduced gene unstable or prone to deletion or rearrangement. The sequence to be expressed may lack suitable restriction sites, start, or stop codons to allow the intended gene to be inserted appropriately into an expression cassette. Apart from the potential adulteration of cloned sequences, the cloning procedure itself can take several days and requires an array of specialist equipment and several lengthy stages of skilled worker input. Because of these limitations, accurate phenotypic analysis of variable gene products may take extended periods of time.

These difficulties are especially significant when analysing the cellular tropism conferred by different HIV-1 Envs derived from primary virus for several reasons. Firstly, the identification of a certain number of non-functional sequences derived from defective virus present *in vivo* would be expected, given the error prone replication of HIV, making deleterious mutations introduced during cloning difficult to separate from genuine *in vivo* sequences. Secondly *env* gene sequences may vary extensively within single infected individuals and the sequence signatures and consequent molecular contortions within Envs leading to differential cellular tropism are incompletely understood. Therefore if important residues were lost or altered during the amplification process it may go unnoticed upon inspection of nucleotide

or amino acid sequence data obtained from clones. Thirdly Env expression is toxic in many eukaryotic cell types and if genes are not fully sequenced prior to phenotypic analysis then apparent failure to express artefactually truncated or otherwise adulterated sequences derived from passage through bacteria may be mistaken for toxicity of *env* gene products in these eukaryotic cells. Use of limiting dilution PCR for amplification of *env* sequences from biological samples avoids many of the artefacts associated with PCR and cloning procedures such as recombination and re-sampling (Chapter 4). However, developing strategies to negate the effects of polymerase induced single nucleotide misincorporations present in clones derived from PCR products is also important for thorough analysis.

Given the limitations of standard cloning techniques outlined above a rapid method for attaching promoter and poly (A) sequences to PCR products that does not require propagation in *E. coli* or other cells before transfection into eukaryotic cells would be extremely useful. One report describes such a procedure (Sykes and Johnston, 1999), whereby a gene, a promoter, and a polyadenylation signal, were separately PCR amplified with primers containing a 12bp to 15bp complementary stretch in which deoxyuridines were incorporated every third position. The PCR products were then treated with uracil-DNA glycosylase to excise uracil residues, therefore providing complementary overhangs suitable for linking the separate PCR products. Resultant DNA was found to be suitable for transfection directly into both mice and cells in culture and resulted in expression of the introduced gene. This approach, while novel, provided fairly poor purity of the final product DNA. We therefore proposed to develop an alternative method for generating this promoter-gene-poly

(A) DNA and tested its efficacy in expression of a variety of genes including HIV-1 *env*.

5.2 Aims

1) To develop a rapid method for the generation of large amounts of coding DNA with the appropriate promoter and poly (A) sequences attached for expression in eukaryotic cells without the use of cloning and plasmid expansion in *E. coli*.

2) Assess the functionality of standard reporter genes when modified in this way by direct transfection into eukaryotic cells followed by analysis of gene expression.

3) To assess and improve the general robustness of this method by using open reading frames of different sizes, and requiring more complex post translational modifications, including HIV-1 *env* amplified at limiting dilution by nested PCR from small samples of biological tissue.

5.3 Materials and methods

As described in chapter 2.

5.4 Results

5.4.1 Assessment of linear PCR product DNA for transient gene expression

The proposed strategy involved the use of PCR and overlap extension reactions to create the appropriate DNA for direct transfection (see below). It has been previously noted that in eukaryotic gene expression experiments supercoiled plasmid DNA is more efficient at inducing transient gene expression, perhaps because DNA in this

conformation interacts with chromatin and is expressed in a similar manner to a cellular gene. In contrast linear DNA may be more recombinogenic and so more suitable for stable gene expression after integration into the host cell genome (Cepko, 2001). In order to assess the efficiency of linear, PCR generated DNA in producing transient gene expression in eukaryotic cells the entire promoter-EGFP-poly (A) segment of pEGFP-C1 was amplified by PCR (Fig 5.1). The resultant 1600 bp amplicon therefore contained the cytomegalovirus immediate early (CMV IE) promoter attached to an enhanced green fluorescent protein (EGFP) open reading frame (ORF) followed by an SV40 poly (A) sequence. This PCR product was purified using a spin column and 1 μ g of the resultant DNA electroporated into 293 T cells. Two days after electroporation the cells were observed under a fluorescent microscope and high levels of gene expression was observed. Negative control electroporation reactions containing either no DNA or dilutions of pEGFP clone equivalent to those that would be present at the end of a PCR reaction were also performed and resulted in no fluorescence (data not shown). This verified that positive results were due to PCR product expression and not expression of small amounts of original template clone carried through the PCR reaction.

5.4.2 Development of PCR expression constructs

Given the efficient gene expression of linear PCR products, a strategy for producing PCR expression constructs was undertaken and tested for various ORFs. The process of ligating a PCR product into a vector is seldom 100% efficient and a certain amount of vector self-ligation occurs even when the 5' phosphate in the reaction is contributed by the PCR product to ensure correct orientation of insert as described here (refer to section 2.8).

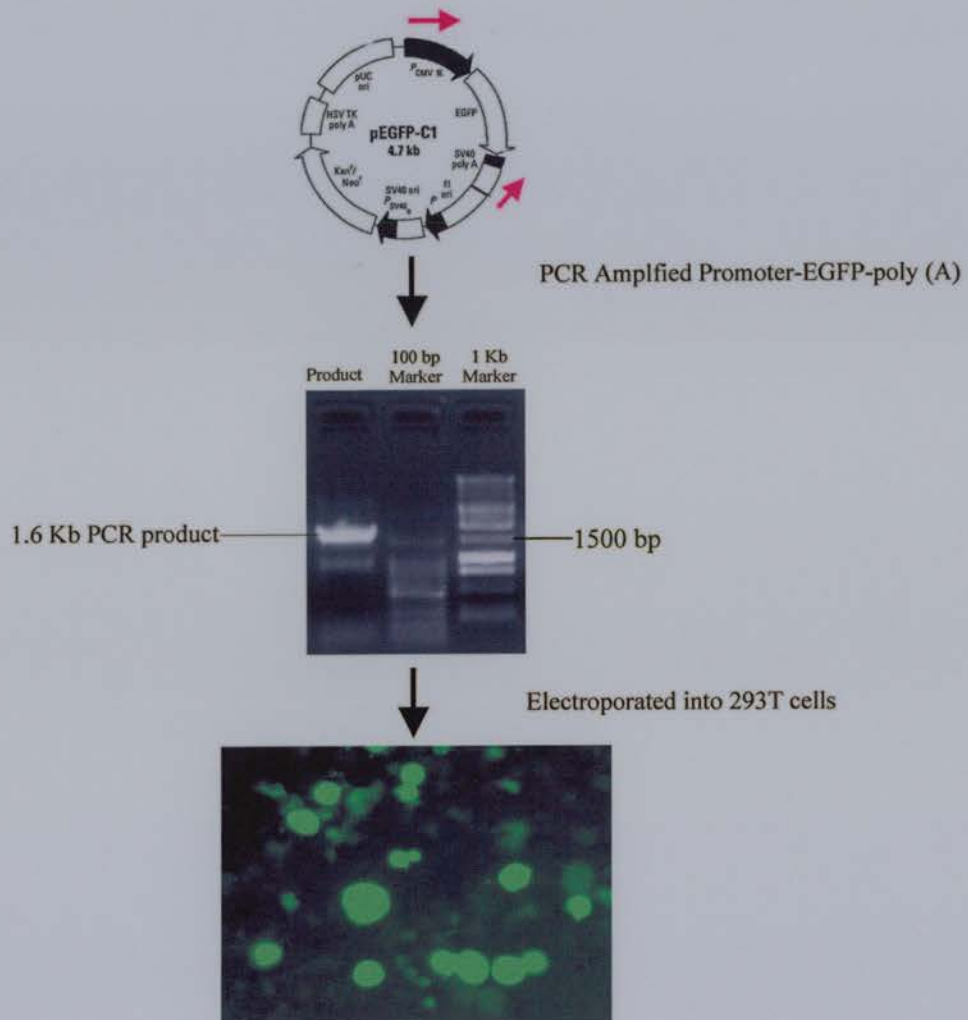


Figure 5.1. Expression of PCR product in 293T cells. The 1600 bp Promoter-EGFP-Poly (A) portion of pEGFP-C1 was amplified by PCR then transfected into 293T cells by electroporation. EGFP protein expression was assessed after 2 days by visualisation of cells under a fluorescent microscope (original magnification: $\times 40$).

Hence if a ligation reaction product is used directly as template in a PCR reaction from the start of the promoter to the end of the Poly (A) signal, spanning the multiple cloning site, the vast majority of resulting PCR product is derived from self-ligated vector. This is due to the amplification process being more efficient in the shorter, self-ligated vector, than in ligation products containing appropriate inserts. The effect becomes more apparent as longer inserts are amplified, and with inserts over approximately 1 Kb no visible product corresponding to the promoter-insert-poly (A) sequence is present (data not shown). Therefore an alternative approach was employed. A schematic diagram of all amplification steps in the process is provided (Fig. 5.2). The validity of this approach was then tested by generating PCR expression constructs using several different genes. Open reading frames were amplified then ligated into pCR3.1 unidirectional vector, an expression vector with the multiple cloning site flanked by the immediate early promoter of CMV and the bovine growth hormone (BGH) poly (A).

These sequences allow the transcription and translation of a cloned gene when large amounts of plasmid DNA are introduced into eukaryotic cells. Generally, in order to generate enough DNA for transfection the plasmid must first be transformed into and amplified in *E. coli* before re-purification. Here the ligated insert was diluted and then two overlapping fragments amplified by PCR. Fragment A consisted of the promoter and start of the gene of interest, and fragment B incorporated the end of the gene along with the poly (A) signal. PCR products A and B were mixed with a DNA polymerase and dNTPs then subjected to thermal cycles allowing extension of the mixed fragments resulting in the production of a single promoter-Gene-poly (A) product.

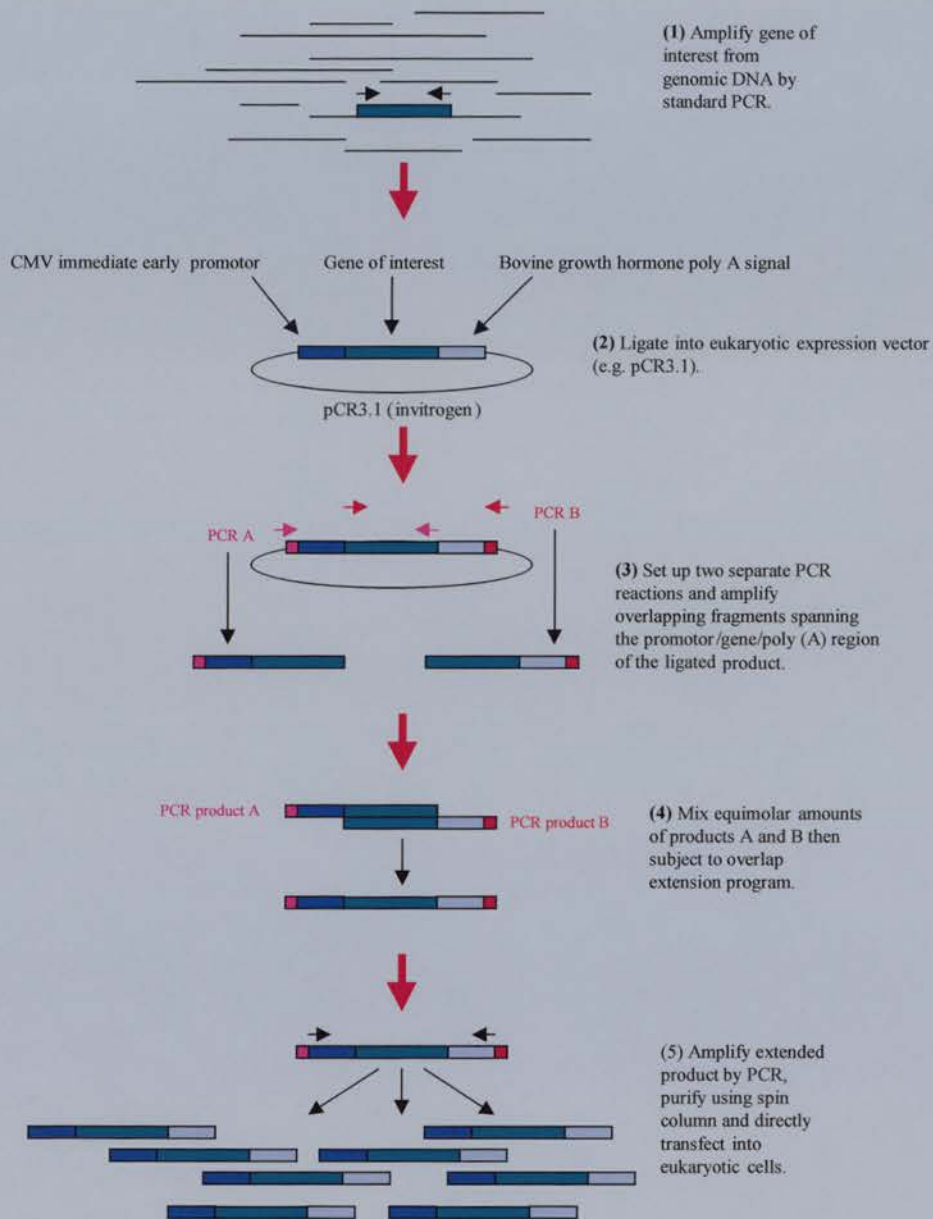


Figure 5.2. Schematic representation of amplification steps used in production of PCR expression constructs.

This template was amplified with primers at the ends and the resultant product then purified and concentrated by standard spin column purification to remove unwanted reagents in the PCR mixture such as primers and detergent. The final products therefore consisted of the gene of interest with the appropriate sequence information attached for expression in eukaryotic cells.

EGFP

The EGFP gene was amplified by single round PCR using the plasmid pEGFP-C1 as template. The antisense PCR primer was modified from vector sequence to incorporate an amber (TAG) stop codon. A PCR expression construct was generated as shown (Fig. 5.3.A) with the same sense and antisense primers used for initial amplification also used in production of PCR products B and A respectively. The final product generated was not ideal (Fig. 5.3.B), with several non-specific bands also present at concentrations nearly equivalent to that of the desired product. However, the purity of this construct was broadly equivalent to that achieved in a report describing a similar approach (Sykes and Johnston, 1999). This construct was then electroporated into 293T cells in parallel with appropriate controls and cells were assessed for EGFP protein production after 48 hours. Half the amount (0.5 μg) of EGFP linear expression construct was used in transfection compared to the pEGFP-C1 positive control to compensate for additional non-coding sequences present in the vector but not in the PCR construct. EGFP expression was observed in both cells electroporated with pEGFP-C1 control plasmid and with the PCR expression construct but absent in cells electroporated with no DNA present (Fig. 5.3.C).

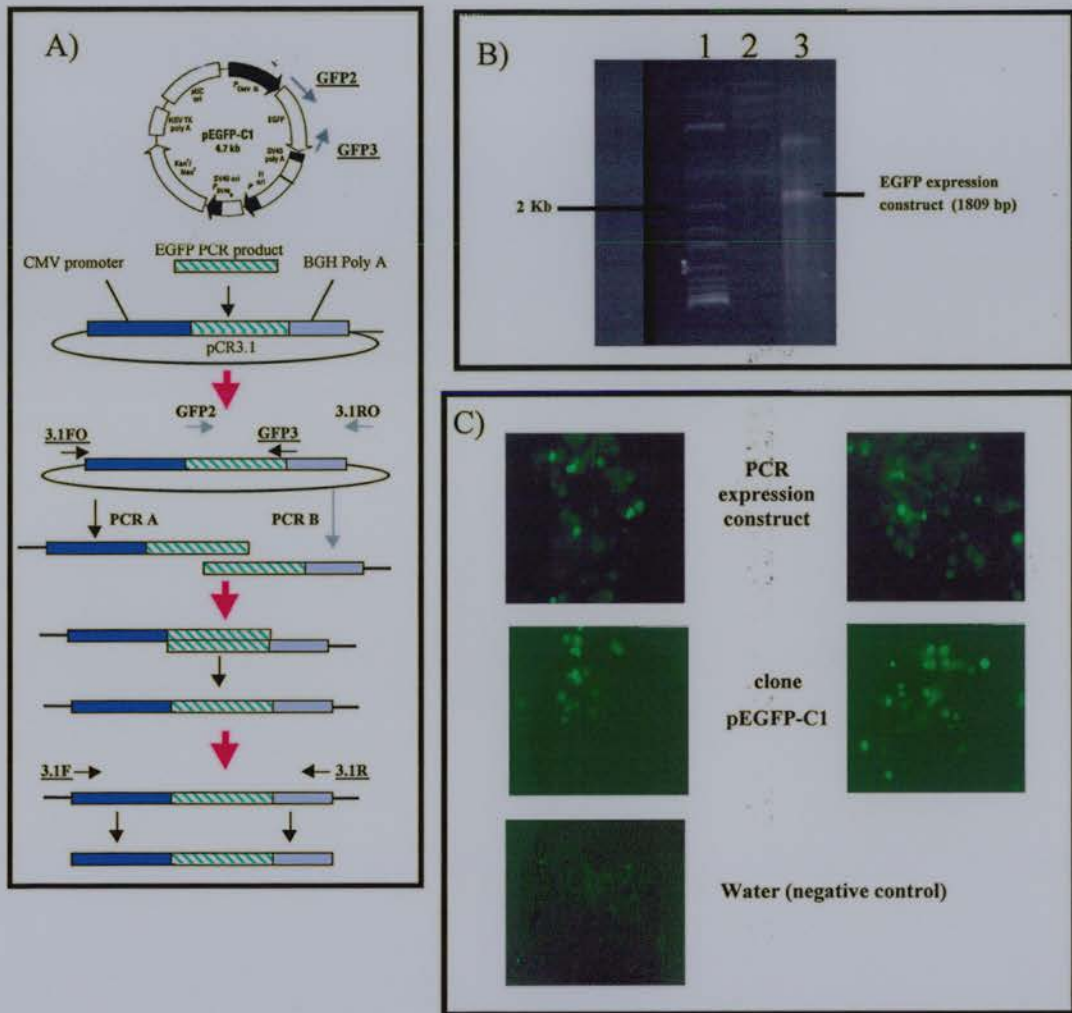


Figure 5.3. EGFP PCR expression construct. A) Schematic representation of primer binding sites and amplification steps, primers are underlined. B) Final EGFP PCR expression construct used for transfection. Lane 1: 1 Kb DNA marker, Lane 2: 100 bp DNA marker, Lane 3: PCR expression construct. C) 293T cells 48 hours after electroporation with EGFP PCR expression construct (PCR product), pEGFP-C1 clone (positive control), or water (negative control). Original magnification: $\times 10$.

Luciferase

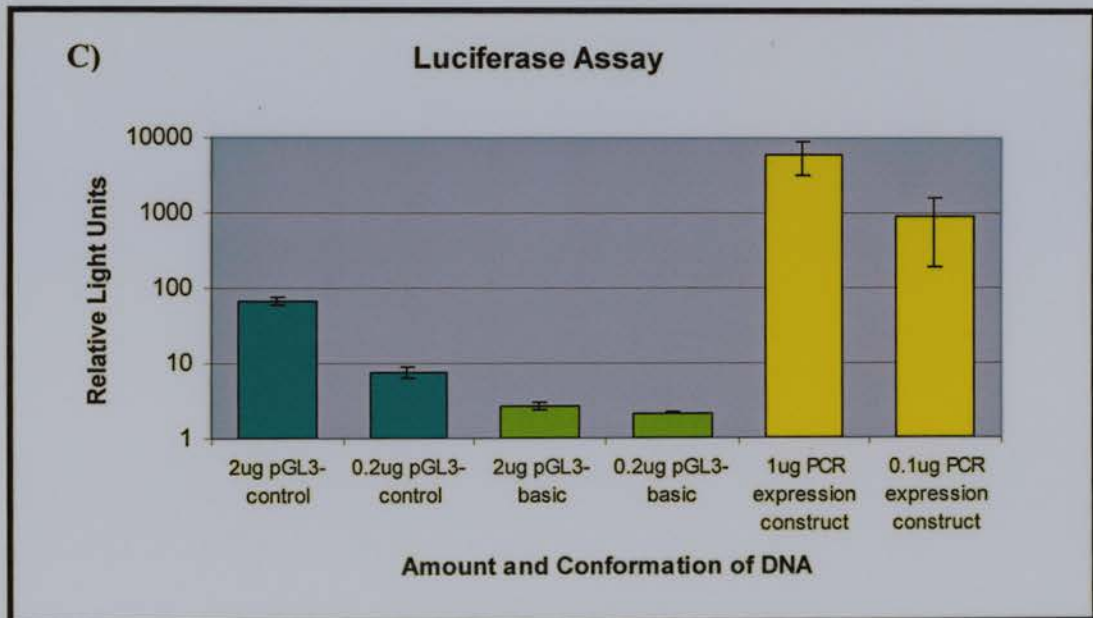
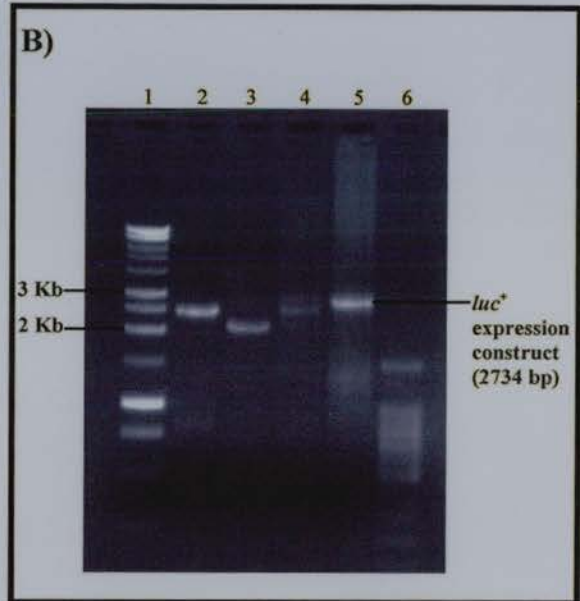
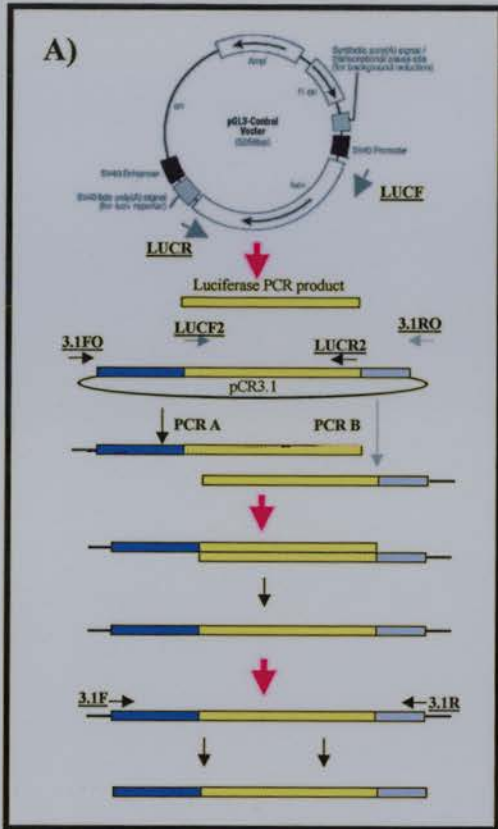
The Firefly luciferase (*luc*⁺) gene was amplified by single round PCR using the pGL3-control (Promega) plasmid as template. To improve the specificity of the amplification process compared with EGFP and to avoid preferential amplification of truncated, ligated PCR products the luciferase PCRs A and B were performed using primers binding 10 bp in from the ends of the insert along with the relevant vector primers (Fig. 5.4.A). The products of PCR A, PCR B, the extension reaction, and the final PCR expression construct are shown (Fig. 5.4.B). The purity of the final product was therefore enhanced by the altered primer positions. This construct was transfected into 293T cells and these cells assayed after 48 hours for luciferase activity. Again half the amount of PCR construct (1 µg) was added compared to positive (pGL3-control) and negative (pGL3-basic) control vectors. The entire process was repeated in three independent experiments. Results of the luciferase assay are shown (Fig. 5.4.C).

For each DNA dilution luciferase activity in cells electroporated with the *luc*⁺ PCR generated expression construct was approximately 100 fold that found in cells electroporated with positive control plasmid.

HIV-1 *nef*

The HIV-1 *nef* gene was amplified by nested PCR from brain tissue obtained at autopsy from the right occipital lobe of an HIV infected individual displaying giant cell encephalitis (NA021). Approximately 600 proviral templates were present in this reaction as determined by PCR using V3 primers.

Figure 5.4. Luciferase PCR expression construct. A) Schematic representation of primer binding sites and amplification steps, primers are underlined. B) PCR products Lane 1: 1 Kb DNA marker, Lane 2: PCR A, Lane 3: PCR B, Lane 4: extension reaction product, Lane 5: final construct, Lane 6: 100 bp DNA marker. C) Luciferase assay result. 293T cells were analysed for luciferase activity 2 days after electroporation with different amounts of positive control clone (pGL3-control), negative control clone containing no promoter or poly (A) signal sequences (pGL3-basic), or *luc*⁺ PCR expression construct. Error bars represent the standard error of three independent experiments where control clones were re-isolated from fresh stocks and PCR expression constructs were generated using the newly isolated positive control vector pGL3-control as template.



In order to determine the robustness of the extension step the *nef* gene insert was amplified using one primer binding 5bp into the gene for PCR A, and one binding near the centre for PCR B, along with the appropriate primers in the vector, thus providing the smaller 400 bp overlap for subsequent extension (Fig. 5.5.A). Nef protein production was assessed using a Nef specific monoclonal antibody for immunocytochemical staining. Staining was evident on 293T cells transfected with 1 µg of the *nef* PCR product but not on negative control cells two days after electroporation (Fig. 5.5.B).

HIV-1 *env*

HIV-1 *env* was amplified by limiting dilution nested PCR using the same DNA sample as for the *nef* gene above. The original *env* PCR product was approximately 3 Kb in size and in order to circumvent potential problems with amplifying a gene of this size along with promoter or poly (A) sequences the PCRs A and B for this gene were each designed to be less than 3 Kb in length but allow a 400bp overlap for extension reaction (Fig. 5.6.A).

The final reaction product consisted of a discrete band at approximately 4 Kb (Fig. 5.6.B). Determining whether a fully functional HIV-1 envelope protein had been produced after transfection into eukaryotic cells required a more complex assay system. This protein undergoes a number of post-translational modifications including cleavage of gp160 into gp120 and gp41 and extensive glycosylation, resulting in the presence of fully functional proteins on the surface of progeny virions. A viral pseudotype assay as described in chapter 4 was therefore undertaken allowing assessment of Env protein functionality when DNA for transfection was generated by production of PCR expression constructs.

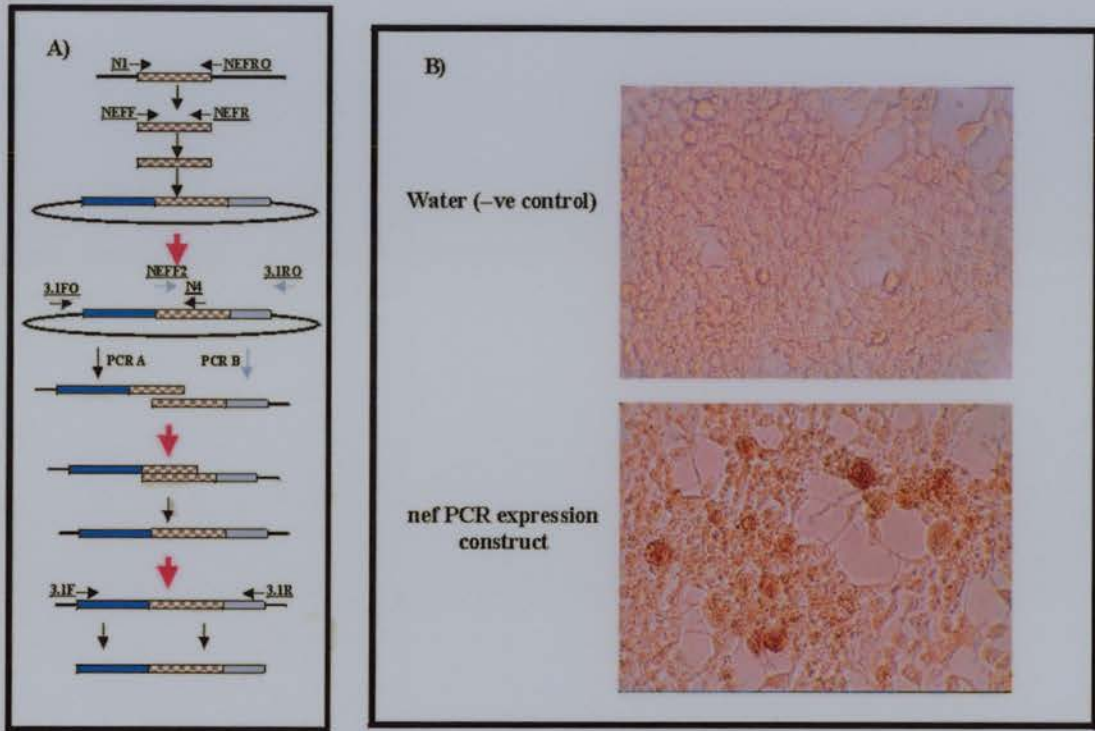


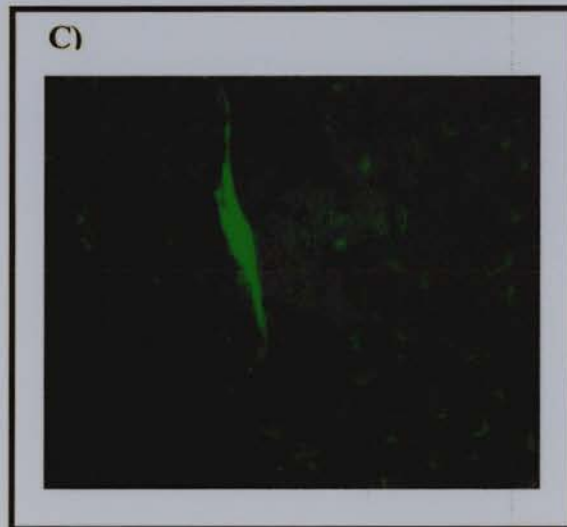
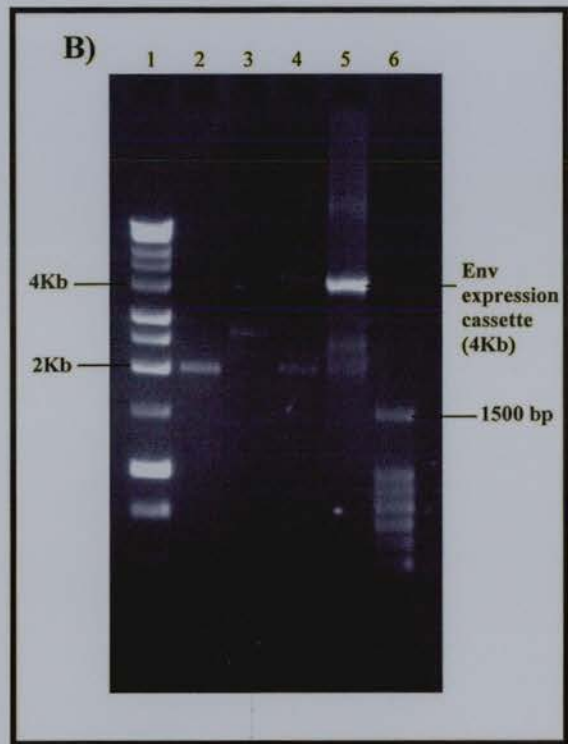
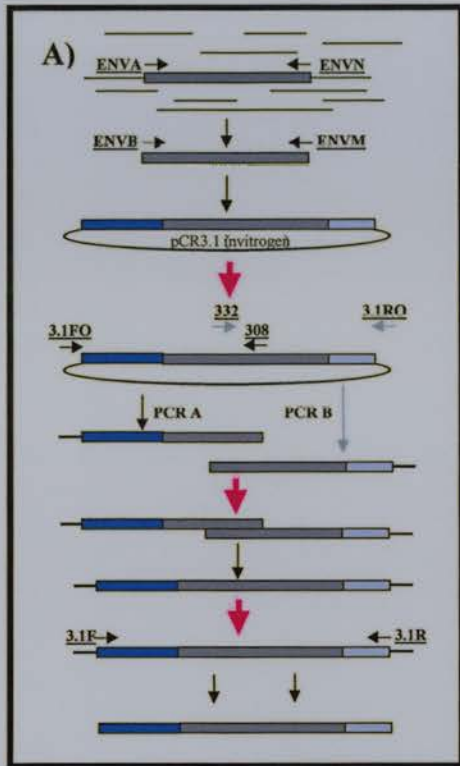
Figure 5.5. Nef expression. A) Schematic representation of primer binding sites and amplification steps, primers are underlined. B) 293T cells were electroporated with water or a *nef* PCR expression construct and stained 48 hours later by immunocytochemistry using Nef specific monoclonal antibodies. Original magnification: $\times 40$.

Figure 5.6. Env expression.

A) Schematic representation of primer binding sites and amplification steps. Primers are underlined.

B) Gel picture of amplification products. Lane 1: 1 Kb DNA marker, Lane 2: PCR A, Lane 3: PCR B, Lane 4: extension reaction product, Lane 5: final construct, Lane 6: 100 bp DNA marker.

C) Infected CD4 and CCR5 expressing NP2 cell viewed under a fluorescent microscope 48 hours after addition of pseudotype virus created with an *env* PCR expression construct. GFP in place of *nef* in the *env* defective proviral backbone is expressed upon infection of a new cell.



Pseudotype virus was produced using envelopes generated with PCR expression constructs and this virus was used for infection of standard indicator cell lines (NP2) expressing the HIV receptors CD4 and CCR5. Pseudotype virus generated with PCR expression constructs was viable, and an example of an NP2 cell expressing CD4 and CCR5, and infected with virus produced in this way is shown (Fig. 5.6.C). A comparison of single round infection assay results using PCR expression construct generated *env* pseudotype virus with that constructed using consensus-cloning procedures (chapter 4) on the same sequences is provided (Table 5.1). It should be noted that these two experiments were performed at different times, and a standardised amount of harvested pseudotype virus was not added to target cells. Therefore the comparison serves only to show that infectious pseudotype virus can be created from identical sequences using both techniques, and does not provide any information on the relative infectivity of virus produced with the different *env* sequences or using the different techniques. Positive control YU2 envelope expressing virus produced approximately 100 to 1000-fold more foci of infection than that generated with patient derived Envs.

This disparity may reflect improved HIV gene expression when driven by the truncated HIV LTR promoter in the pSVIII vector as compared with the CMV IE promoter in pCR3.1.

Sample*	Clone	PCR Expression construct
RO1	2.1×10^3	1.2×10^3
RO3	1.75×10^3	6.5×10^2
RO6	4.1×10^2	2.8×10^2
YU2 control	9.5×10^4	$>1.5 \times 10^5$
No env (pNL4.3 Δ envGFP only)	0	0

Table 5.1. Infectious titre of pseudotype virus on NP2 cells expressing CD4 and CCR5. Comparison of virus generated with *env* consensus clones (chapter 4) and PCR generated expression constructs made from the same sequences. Numbers correspond to focal forming units/ml. * RO: limiting dilution *env* PCR products as described in chapter 4, YU2: control clone from CCR5-using HIV-1 isolate.

5.5 Discussion

The procedures described here present a rapid, simple, method for the generation of large amounts of an open reading frame (ORF) flanked by the promoter and poly (A) sequences necessary for expression in eukaryotic cells without the need for prior amplification in bacterial cells. Initial transfection experiments using a promoter-EGFP-poly (A) fragment amplified as a single PCR product from a control plasmid demonstrated that PCR generated, linear DNA was suitable for gene expression in eukaryotic cells (Fig. 5.1). Consequently we designed a technique for generation of such an expression construct directly from ligation reaction products by PCR and overlap extension procedures (Fig. 5.2). An EGFP PCR expression construct was then generated by this method (Fig. 5.3) and green fluorescent protein expression after electroporation into 293 T cells compared with that of the positive control clone pEGFP-C1 (Fig. 5.3.C). This illustrated that sequences modified and amplified in this way were suitable for producing gene expression in eukaryotic cells at similar levels to the parental positive control clone. Amplification of *luc*⁺ PCRs A and B using primers that bind 10 bp inside the original PCR product improved the specificity of the amplification process and therefore the purity of the final product (Fig. 5.4). Luciferase results provided quantitative data on gene expression, with 100-fold activity detected in cells transfected with *luc*⁺ PCR expression construct when compared with cells transfected with the pGL3-control positive control clone. This large difference in expression levels was consistent in three completely independent experiments and is likely to reflect improved activity of the CMV IE promoter in 293T cells when compared with the SV40 promoter present in the pGL3-control plasmid. Possibly the altered DNA conformation (linear PCR product DNA

compared with covalently closed circular supercoiled plasmid DNA) was also a factor in the differing level of gene expression. In any case large amounts of luciferase activity were detected, demonstrating high levels of protein production in transfected cells. These results also demonstrated that this approach could be used for expression of the 1649 bp *luc*⁺ gene as well as the shorter 723 bp EGFP gene.

To assess the efficacy of using this procedure for analysis of HIV-1 gene products the HIV-1 *nef* and *env* genes were amplified and tested using a similar process. Amplification of the HIV-1 *nef* gene by nested PCR using template DNA extracted from autopsy tissue followed by production of an expression construct established that the procedure was indeed suitable for use with genes amplified from small amounts of biological tissue. Generation of this construct was undertaken using a primer for PCR A that binds near the centre of the gene, indicating that a relatively small overlap of 400 bp between PCRs A and B is sufficient for an efficient subsequent extension reaction. Clearly the use of DNA containing several hundred original templates may make this reaction subject to recombination during PCR. However, this would not be significant if the process were to be used on non-variable genes. Also, in some cases an average phenotype from a population of sequences may be all that is required, or a quick screen of a large selection of variable genes to identify the presence of some with a desired property. Nef protein expression was observed by immunocytochemistry using a Nef specific monoclonal antibody in cells electroporated with a *nef* PCR expression construct but not in negative control cells electroporated in the absence of DNA (Fig. 5.5). Finally, production of functional pseudotype virus using an *env* PCR expression construct demonstrated that the system could be used to produce a fully functional viral envelope protein requiring

complex post translational modifications using DNA amplified from biological samples. Therefore the procedure is suitable for use in phenotypic analysis of HIV-1 Env proteins without the possibility of alterations in *E. coli*. Moreover when used in conjunction with original PCR amplification at limiting dilution then any deleterious effects of polymerase misincorporation during PCR are negated (the rationale for this is as for generation of consensus clones in chapter 4). One possibly important issue in the method described for amplification of *nef* and *env* PCR expression constructs is that some of the primers used were specific for areas in the centre of the original PCR amplicon. Primers placed in this way may lead to the formation of chimeric molecules if genetically distinct sequences are present at the end of a PCR reaction, such as probably would be the case in the *nef* PCR reaction described. Also, even if the original PCR is amplified at limiting dilution as for the *env* gene here then the central primer sequences used for amplification of PCRs A and B would be incorporated into the final construct. The primers used here were designed to correspond to fairly conserved regions of *env*, and these experiments intended only as proof of principle. Nevertheless in future a more rigorous approach would be to use primers for PCR A and B that are only 5 or 10 bp from the ends of the original amplicon as described for *nef* PCR A the *luc*⁺ gene.

In addition to providing a means for expressing genes in eukaryotic cells that are toxic in *E. coli* and thus may be refractory to standard cloning procedures, the method described here has a number of other advantages over more established techniques. DNA suitable for transfection can be produced in a single working day. Also, the specialist worker knowledge and array of equipment needed for bacterial cloning procedures are not required. If variants of the same gene are to be

phenotypically analysed then a single promoter and poly (A) pair, designed with appropriate extensions corresponding to the ends of the gene could be used for generation of expression constructs for all variants of that gene, thereby negating the need for ligation into a vector and amplification of overlapping fragments before overlap extension. One obvious potential application of this technique is in the phenotypic analysis of resistance to antiviral agents.

HIV-1 entry inhibitors including those that specifically target HIV-1 Env are currently under development, and should soon be available for treatment of infected individuals (D'Souza et al., 2000). In clinical trials, use of the fusion inhibitor T20 effectively reduced plasma HIV RNA in infected individuals (Kilby et al., 1998). However, an escape mutant to a fusion inhibitor has been identified in tissue culture (Rimsky et al., 1998), and it would be expected for similar resistance mutations to arise in virus within treated patients. As more elaborate cocktails of antiviral agents are targeted towards Env, it is probable that mutations in *env* that confer drug resistance will become increasingly complex, as has been the case with other HIV-1 protein targets for antiviral therapy (Jacobsen et al, 1995; Winters et al, 1998; Larder and Kemp, 1989; Richman et al, 1994). It may therefore be expected for phenotypic assays to be required for the complete characterisation of drug resistant viral phenotypes. The rapid generation of envelope sequences suitable for phenotypic analysis without the requirement for cloning as described here may prove useful in the context of these future applications.

Chapter 6: Final discussion

The genotypic and phenotypic characteristics of HIV-1 within the brain remain incompletely defined. Many factors could lead to the apparent tissue specific compartmentalisation often observed in individuals with AIDS and HIVE (Epstein et al, 1991b; Epstein et al, 1991a; Haggerty and Stevenson, 1991; Korber et al, 1994a; Pang et al, 1991; Chang et al, 1998; Donaldson et al, 1994a; Gartner et al, 1997; Gorry et al, 2001; Hughes et al, 1997a; Shapshak et al, 1999). Similarly, evidence for the role of recombination between strains that have diverged within a single infected individual is currently confined to a limited number of studies (Jung et al, 2002; Dykes et al, 2000). One aim of this project was to search for evidence of recombination between the apparent tissue-specific variants in the brain and contemporary HIV-1 sequences present in the lymphoid system of the same individuals, thus providing information on both of these important issues.

Factors that could in principal lead to the development of a separate virus population in the brain include physical isolation of the two populations and differences in virus turnover rates in the two compartments. Alternatively, the genetic segregation may be a result of adaptation to the particular microenvironment and available host cells within the brain. If this is the case then lymphoid variants may enter the brain with relative ease, but be poorly equipped for prolonged survival in the unique environment of the CNS. These newly infiltrating lymphoid variants could circumvent such a restriction by recombining with virus already *in situ*, and acquiring the necessary sequences for neuroadaptation while retaining “lymphoid-like” sequence segments elsewhere in their genomes.

To address these possibilities, and obtain a broad overview of the sequence relationships manifest in different areas of the viral genome in brain and lymphoid tissue, consensus sequences (and in one study subject representative clones) were obtained corresponding to the p17^{gag} and V3 HIV-1 subgenomic regions. These sequences were recovered from several different areas of brain and the lymphoid compartment of four study subjects with evidence of HIVE upon autopsy. Synonymous mutations in the p17^{gag} region had previously been demonstrated to accumulate at a relatively fixed rate over time, and could therefore be used as a molecular clock (Hughes et al, 1997b; Wang et al, 2001; Holmes et al, 1993; Kasper et al, 1995). In two study subjects the lineage in p17^{gag} that contained lymphoid sequences was separate from most brain-derived p17^{gag} sequences, allowing mean synonymous pairwise J-C distances to be calculated, and the approximate divergence time of the two populations to be estimated. In two individuals this divergence time preceded the onset of AIDS, perhaps suggesting a pre-AIDS entry of HIV-1 into the CNS of these study subjects.

Further information was obtained from the observed phylogenetic relationships between viral variants present in different anatomical sites. In an asexually reproducing population where recombination has not occurred, one would expect for the sequence relationships between different isolates to be the same regardless of the genomic region analysed. This assumption is of course subject to the requirement that the genomic regions in question must contain a sufficient number of nucleotide differences to be phylogenetically informative. In each study subject, the relationships between proviral sequences from different anatomical sites differed depending upon whether V3 or p17^{gag} sequences were being compared. The most

obvious explanation for these discordant phylogenetic relationships is the occurrence of recombination at a point in the viral genome between the p17^{gag} and V3 regions. An association was also noted between the sequence divergence in V3 both within brain and between brain and lymphoid virus, and the severity of HIV. This may reflect increased genetic drift in V3 sequences within the brain, permitted by a weaker anti-HIV humoral immune response in the CNS.

In order to further characterise the role of recombination during HIV-1 infection of the brain it was necessary to obtain longer sequences from single proviruses present in different anatomical sites. The *env* gene was chosen for this application, as it was most likely to contain the determinants of differential HIV-1 cellular tropism, and was of sufficient length to potentially contain both “brain-like” and “lymphoid-like” segments of sequence. The standard method for obtaining sequences such as this is by PCR amplification of a large number of template molecules in a single reaction, followed by molecular cloning of the PCR products. A number of experiments designed to assess the validity of these techniques demonstrated that they were unsuitable for the purpose of obtaining sequences to study *in vivo* recombination. For example recombination during PCR amplification of an approximately 3 Kb genomic region encompassing the *env* gene was found to occur at a very high level, and would make the task of identifying natural recombinants impossible. Hence, *env* genes for the detection of recombinants were all amplified by limiting dilution PCR. Sequences were obtained from lymph node 2, and both the right occipital and left temporal brain regions of study subject NA021. While it is difficult to unequivocally identify recombinants unless the genotype of the parental sequences is known, the complete sequences used in this study did cluster broadly into phylogenetically

distinct groups. It was therefore possible to use a sliding window approach to compare the phylogenetic affinity of 500 bp segments along the length of a single sequence with all the remaining sequences in the three groups. If the phylogenetic affinity switched from one group to another, then that sequence was deemed likely to be a recombinant. In this way several *env* recombinants were identified, where in one area of the *env* gene they were almost indistinguishable from lymphoid variants, and in another they were closely related to variants in one of the regions of brain. Interestingly, no sequence from brain grouped with lymphoid sequences in the V1-V3 region of *env*, perhaps reflecting the importance of this subgenomic region as a determinant of HIV-1 cellular/tissue tropism (Chesebro et al, 1991; Hwang et al, 1991; Fouchier et al, 1992; Speck et al, 1997; Carrillo and Ratner, 1996; Shioda et al, 1991; Martin et al, 2001). In several cases, switches in phylogenetic affinity from members of one brain region to members of the second brain region was also evident in different segments of the same sequence, suggesting recombination between viral variants present in anatomically distinct areas of brain. Again, if viral variants from different sites can recombine with one another, then they must have at some point infected the same cell. Therefore these results suggest some spreading or movement of HIV within the brain, allowing variants that had diverged previously to come into close proximity and recombine.

Together these results have several implications. Firstly, it appears that recombination between variants that have diverged within a single infection is a common event. This high level of recombination is perhaps to be expected, given the global prevalence of recombinant HIV genomes (Liitsola et al, 1998; Gao et al, 1998; Liitsola et al, 2000; Montavon et al, 1999). However only a limited number of

studies have attempted to identify recombinants between closely related variants as described here (Brown and Cleland, 1996; Jung et al, 2002). Recombination such as this may bring together separate constellations of mutations associated with drug resistance, allowing the outgrowth of resistant virus much more rapidly than if each had to be acquired by the step-by-step accumulation of single mutations. In the context of HIV-1 and the brain, the identification of brain/lymphoid recombinants suggests that the two compartments may not be as physically separated as previously supposed. If lymphoid variants are able to access the brain and recombine with virus already present, then it appears unlikely that the genetic compartmentalisation often seen between brain and lymphoid variants could be due to physical isolation or differences in viral turnover rates in the two compartments. Therefore regions of the viral genome that do demonstrate tissue specific groupings (generally the V1-V3 regions of *env*) are likely to be genetically distinct for a reason. If there were no selective advantage associated with these sequences in the brain then it would be expected for them to be replaced by, or mixed with, the lymphoid variants that appear to be able to enter and exchange other subgenomic regions by recombination. Recombination between the two populations could also allow virus in the brain to acquire advantageous mutations such as cytotoxic T lymphocyte escape mutations, or antiviral resistance mutations, while still retaining their neuroadapted phenotype. The reverse could also be true, with the poor penetration of many antiretroviral agents into the CNS potentially providing an optimal environment for the generation HIV-1 variants with resistance to antiviral drugs (Lewis et al, 1996; Yarchoan et al, 1988). It has been suggested that a neuroadapted HIV-1 *env* phenotype may involve an increase in coreceptor affinity at the expense of a higher susceptibility to antibody

neutralisation (Gorry et al, 2002a; Martin et al, 2001). Thus a neuroadapted viral phenotype may be poorly adapted for replication in the lymphoid system. In this case drug resistance mutations that have evolved in the brain could potentially be acquired by lymphoid variants via recombination and carried back to the lymphoid system.

Mutations in *env* that affect receptor interactions, cellular tropism, and potentially tissue tropism are often highly complex and strain dependent. Apart from the frequent, but not complete, association between particular amino acid residues in V3 and CXCR4 usage (Speck et al, 1997) it is not possible to predict the phenotype of a particular Env from the primary nucleotide or amino acid sequence. This is especially significant in the context of HIV-1 in the brain, where potential neuroadaptation and neurovirulence does not appear to be due to a change in coreceptor usage. The limited numbers of comprehensive studies into interactions between brain-derived Envs with HIV-1 cellular receptors suggest that determinants of neurotropism may be subtle. For example they may include increases in receptor affinity and alterations in the specific manner that coreceptors are used, which are conferred by mutations in a complex and strain specific manner (Gorry et al, 2002a; Martin et al, 2001). These caveats make it of paramount importance that HIV-1 *env* sequences analysed for particular phenotypic properties accurately represent those present *in vivo*. A related issue is that technological advances have led to increasing numbers and lengths of nucleotide sequences being obtained from viral variants in HIV-1 infected individuals. Sequence analysis of HIV-1 variants in the context of a single infection can provide many insights into HIV biology. However, it is important that sequences obtained are free from *in vitro* artefacts.

A number of potential *in vitro* artefacts are associated with common PCR and cloning methods for obtaining sequences from heterogeneous viral populations. These include the incorporation of single nucleotide polymorphisms during PCR, recombination during PCR, resampling of clones, and alterations of sequences during plasmid expansion in bacteria. The relative importance of each of these artefacts was assessed in the context of the *env* gene. Results showed that all of these *in vitro* artifacts had a significant impact upon the final sequences obtained. It may therefore be expected (if there were no specific attempts to avoid such artefacts) that this type of sequence adulteration or misrepresentation of sequence diversity would be present in other studies that used similar techniques. Of particular interest was the observed resampling of clones, despite the presence of sufficient amplifiable template molecules. Recombination during PCR was also found to occur at a high level, even when only a small number of templates copies from each original genotype were present. Consequently, the fairly common procedure of reducing template copy number to reduce recombination during PCR may not be sufficient in many cases. Thus when the same primer pairs are used amplify mixtures of different template molecules, the level of *in vitro* recombination should be empirically determined for each individual PCR. Inclusion of point mutations during PCR amplification could also alter phenotypic properties of amplified sequences, and may partially obscure clone resampling when it has occurred by artificially increasing the observed nucleotide differences between individual clones. Together these results imply that in order to ensure that amplified sequences correspond to those present *in vivo*, limiting dilution PCR should be used.

The issues described above, together with the toxic effect of lentiviral *env* sequences within *E. coli* observed in this and other studies (Cunningham et al, 1993; Peden, 1992; Wang and Mullins, 1995), suggested that *env* sequences to be used for phenotypic analysis should ideally be obtained by limiting dilution PCR, and not be subjected to passage through bacteria during cloning. To this end a PCR and overlap extension method was developed, where promoter and poly (A) sequences could be attached to an open reading frame without the need for standard cloning and passage through bacteria. This method was found to be effective for expression of the EGFP and *luc*⁺ reporter genes. Further assessment of this method using the HIV-1 genes *nef* and *env* derived from an infected individual also led to efficient expression. The expression of *env* open reading frames was also performed using PCR products obtained at limiting dilution, and therefore conformed to all the ideal criteria for maintenance of *in vivo* sequence characteristics. This method may therefore prove useful for analysis of the subtle determinants in Env that confer differential cell and potentially tissue tropism. Moreover antiviral therapies targeted towards HIV-1 Env should become widely available in the near future (D'Souza et al, 2000; Chan et al., 1998; Kilby et al, 1998). Viral escape mutants to one of these agents: DP178 (Wild et al., 1994), have been identified in tissue culture experiments (Rimsky et al, 1998). Given the complex collections of mutations that correlate with HIV resistance to currently available therapies (Jacobsen et al, 1995; Winters et al, 1998; Larder and Kemp, 1989; Richman et al, 1994), it may be expected that mutations conferring resistance to antiviral agents targeted against Env proteins will be similarly difficult to interpret at a genotypic level. Hence phenotypic assays will likely be required, designed for testing the susceptibility of patient-derived Envs to different antiviral

agents. The rapid and cloning-free method described here for producing *env* genes suitable for expression in mammalian cells may therefore also prove useful in the context of these potential future applications.

In summary, the results of this investigation suggest that recombination between HIV-1 variants that have diverged during the course of a single infection is a common event. This recombination was readily evident between brain and lymphoid derived HIV-1 variants in several infected individuals with HIV. Therefore it appears that variants from outwith the brain interact with those present in the CNS with some regularity. Consequently it may be inferred that the regions of the viral genome that do show tissue specific segregation probably confer some level of selective advantage for replication in the CNS. The results of investigations into *in vitro* artefacts incorporated during PCR and cloning strongly suggest that limiting dilution PCR should be used for obtaining sequences from heterogeneous viral populations. A method for avoiding each of these artefacts was also presented and may prove useful in future investigations.

Future work

The results of *in vivo* recombination within the *env* gene between HIV-1 variants present in the brain and lymphoid system, while interesting, are based upon virus present in one infected individual. Clearly it is not possible to draw broad conclusions from results derived from a single study subject. Consequently Takumi Shirafuji, a colleague in the laboratory, is currently obtaining viral sequences from a second study subject for use in the same type of analysis. Preliminary results of this analysis are consistent with those presented herein. Kirsty Newman, another coworker in the laboratory, is using a modified version of the overlap extension technique, designed in this case to generate entire HIV-1 proviral sequences with patient-derived envelopes.

Bibliography

- Abebe, A., Demissie, D., Goudsmit, J., Brouwer, M., Kuiken, C. L., Pollakis, G., Schuitemaker, H., Fontanet, A. L., and de Wit, T. F. R. (1999). HIV-1 subtype C syncytium- and non-syncytium-inducing phenotypes and coreceptor usage among Ethiopian patients with AIDS. *AIDS* **13**, 1305-1311.
- Achong, B. G., Mansell, P. W., Epstein, M. A., and Clifford, P. (1971). An unusual virus in cultures from a human nasopharyngeal carcinoma. *J.Natl.Cancer Inst.* **46**, 299-307.
- Adamson, D. C., Kopnisky, K. L., Dawson, T. M., and Dawson, V. L. (1999). Mechanisms and structural determinants of HIV-1 coat protein, gp41- induced neurotoxicity. *J.Neurosci.* **19**, 64-71.
- Aiken, C., Konner, J., Landau, N. R., Lenburg, M. E., and Trono, D. (1994). Nef induces CD4 endocytosis - requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain. *Cell* **76**, 853-864.
- Akroyd, J., Fincham, V. J., Green, A. R., Levantis, P., Searle, S., and Wyke, J. A. (1987). Transcription of Rous sarcoma proviruses in rat cells is determined by chromosomal position effects that fluctuate and can operate over long distances. *Oncogene* **1**, 347-354.
- Albright, A. V., Shieh, J. T., Itoh, T., Lee, B., Pleasure, D., O'Connor, M. J., Doms, R. W., and Gonzalez-Scarano, F. (1999). Microglia express CCR5, CXCR4, and CCR3, but of these, CCR5 is the principal coreceptor for human immunodeficiency virus type 1 dementia isolates. *J.Virol.* **73**, 205-213.
- Alizon, M., Wain Hobson, S., Montagnier, L., and Sonigo, P. (1986). Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. *Cell* **46**, 63-74.
- Alkhatib, G., Broder, C. C., and Berger, E. A. (1996a). Cell type-specific fusion cofactors determine human immunodeficiency virus type 1 tropism for T-cell lines versus primary macrophages. *Journal of Virology* **70**, 5487-5494.
- Alkhatib, G., Combadiere, C., Broder, C. C., Feng, Y., Kennedy, P. E., Murphy, P. M., and Berger, E. A. (1996b). CC CKR5: A RANTES, MIP-1 alpha, MIP-1 beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* **272**, 1955-1958.
- Alkhatib, G., Liao, F., Berger, E. A., Farber, J. M., and Peden, K. W. (1997). A new SIV co-receptor, STRL33. *Nature* **388**, 238.
- Alonso, A. and Peterlin, B. M. (1999). HIV-1: Control of gene expression by the viral regulatory proteins Tat and Rev. In "HIV and the New Viruses" (A. G. Dalgleish and R. A. Weiss, Eds.), pp. 1-28. Academic Press, London.
- An, S. F., Giometto, B., and Scaravilli, F. (1996). HIV-1 DNA in brains in AIDS and pre-AIDS: correlation with the stage of disease. *Ann.Neurol.* **40**, 611-617.
- Anderson, J. A., Teufel, R. J., Yin, P. D., and Hu, W. S. (1998). Correlated template-switching events during minus-strand DNA synthesis: a mechanism for high negative interference during retroviral recombination. *J Virol* **72**, 1186-1194.
- Aramori, I., Zhang, J., Ferguson, S. S. G., Bieniasz, P. D., Cullen, B. R., and Caron, M. G. (1997). Molecular mechanism of desensitization of the chemokine receptor CCR-5: receptor signaling and internalization are dissociable from its role as an HIV-1 co-receptor. *EMBO J* **16**, 4606-4616.
- Arendrup, M., Nielsen, C., Hansen, J. E., Pedersen, C., Mathiesen, L., and Nielsen, J. O. (1992). Autologous HIV-1 neutralizing antibodies: emergence of neutralization-resistant escape virus and

- subsequent development of escape virus neutralizing antibodies. *J.Acquir.Immune.Defic.Syindr.* **5**, 303-307.
- Arion, D., Kaushik, N., McCormick, S., Borkow, G., and Parniak, M. A. (1998). Phenotypic mechanism of HIV-1 resistance to 3'-azido-3'-deoxythymidine (AZT): increased polymerization processivity and enhanced sensitivity to pyrophosphate of the mutant viral reverse transcriptase. *Biochemistry* **37**, 15908-15917.
- Arthos, J., Rubbert, A., Rabin, R. L., Cicala, C., Machado, E., Wildt, K., Hanbach, M., Steenbeke, T. D., Swofford, R., Farber, J. M., and Fauci, A. S. (2000). CCR5 signal transduction in macrophages by human immunodeficiency virus and simian immunodeficiency virus envelopes. *J.Virol.* **74**, 6418-6424.
- Arya, S. K., Guo, C., Josephs, S. F., and Wong Staal, F. (1985). Trans-activator gene of human T-lymphotropic virus type III (HTLV-III). *Science.* **229**, 69-73.
- Asjo, B., Albert, J., Karlsson, A., Morfeldt Manson, L., Biberfeld, G., Lidman, K., and Fenyo, E. M. (1986). Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection. *Lancet* **ii**, 660-662.
- Ayyavoo, V., Mahboubi, A., Mahalingam, S., Ramalingam, R., Kudchodkar, S., Williams, W. V., Green, D. R., and Weiner, D. B. (1997). HIV-1 Vpr suppresses immune activation and apoptosis through regulation of nuclear factor kappa B. *Nature Med* **3**, 1117-1123.
- Bachand, F., Yao, X. J., Hrimech, M., Rougeau, N., and Cohen, E. A. (1999). Incorporation of Vpr into human immunodeficiency virus type 1 requires a direct interaction with the p6 domain of the p55 gag precursor. *Journal of Biological Chemistry* **274**, 9083-9091.
- Bagasra, O., Lavi, E., Bobroski, L., Khalili, K., Pestaner, J. P., Tawadros, R., and Pomerantz, R. J. (1996). Cellular reservoirs of HIV-1 in the central nervous system of infected individuals: Identification by the combination of in situ polymerase chain reaction and immunohistochemistry. *AIDS* **10**, 573-585.
- Baggiolini, M. (1998). Chemokines and leukocyte traffic. *Nature* **392**, 565-568.
- Ball, J. K., Holmes, E. C., Whitwell, H., and Desselberger, U. (1994). Genomic variation of human immunodeficiency virus type 1 (HIV-1) - molecular analyses of HIV-1 in sequential blood samples and various organs obtained at autopsy. *J.Gen.Virol.* **75**, 867-879.
- Balliet, J. W., Kolson, D. L., Eiger, G., Kim, F. M., Mcgann, K. A., Srinivasan, A., and Collman, R. (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.
- Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**, 1209-1211.
- Baribaud, F., Edwards, T. G., Sharron, M., BreLOT, A., Heveker, N., Price, K., Mortari, F., Alizon, M., Tsang, M., and Doms, R. W. (2001). Antigenically distinct conformations of CXCR4. *J.Virol.* **75**, 8957-8967.
- Barin, F., M'Boup, E. M., Denis, F., Kanki, P., Allan, J. C., Lee, T. H., and Essex, M. (1985). Serological evidence for virus related to simian-T-lymphotropic retrovirus III in residents of West Africa. *Lancet* **ii**, 1387-1389.
- Barnes, W. M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc.Natl.Acad.Sci.U.S.A* **91**, 2216-2220.

Barre Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler Blin, C., Vezinet Brun, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*. **220**, 868-871.

Bazan, H. A., Alkhatib, G., Broder, C. C., and Berger, E. A. (1998). Patterns of CCR5, CXCR4, and CCR3 usage by envelope glycoproteins from human immunodeficiency virus type 1 primary isolates. *J Virol* **72**, 4485-4491.

Bebenek, K., Abbotts, J., Roberts, J. D., Wilson, S. H., and Kunkel, T. A. (1989). Specificity and mechanisms of error prone replication by Human Immunodeficiency-1 reverse transcriptase. *J Biol Chem* **264**, 16948-19656.

Bell, J. E. (1998). The neuropathology of adult HIV infection. *Rev.Neurol. (Paris)* **154**, 816-829.

Bell, J. E., Brettle, R. P., Chiswick, A., and Simmonds, P. (1998). HIV encephalitis, proviral load and dementia in drug users and homosexuals with AIDS. Effect of neocortical involvement. *Brain* **121**, 2043-2052.

Bell, J. E., Busuttil, A., Ironside, J. W., Rebus, S., Donaldson, Y. K., Simmonds, P., and Peutherer, J. F. (1993). Human immunodeficiency virus and the brain - investigation of virus load and neuropathologic changes in pre-AIDS subjects. *Journal of Infectious Diseases* **168**, 818-824.

Benichou, S., Bomsel, M., Bodeus, M., Durand, H., Doute, M., Letourneur, F., Camonis, J., and Benarous, R. (1994). Physical interaction of the HIV-1 nef protein with beta-COP, a component of non-clathrin-coated vesicles essential for membrane traffic. *Journal of Biological Chemistry* **269**, 30073-30076.

Benkirane, M., Jin, D. Y., Chun, R. F., Koup, R. A., and Jeang, K. T. (1997). Mechanism of transdominant inhibition of CCR5-mediated HIV-1 infection by ccr5 Delta 32. *J Biol Chem* **272**, 30603-30606.

Bennett, R. P., Nelle, T. D., and Wills, J. W. (1993). Functional chimeras of the rous sarcoma virus and human immunodeficiency virus gag proteins. *J.Virol.* **67**, 6487-6498.

Benson, R. E., Sanfridson, A., Ottinger, J. S., Doyle, C., and Cullen, B. R. (1993). Downregulation of cell-surface CD4 expression by simian immunodeficiency virus nef prevents viral super infection. *J.Exp.Med.* **177**, 1561-1566.

Berger, E. A. (1997). HIV entry and tropism: the chemokine receptor connection. *AIDS* **11**, S3-S16.

Berger, E. A., Doms, R. W., Fenyo, E. M., Korber, B. T. M., Littman, D. R., Moore, J. P., Sattentau, Q. J., Schuitemaker, H., Sodroski, J., and Weiss, R. A. (1998). A new classification for HIV-1. *Nature* **391**, 240.

Bernacchi, S., Stoylov, S., Piemont, E., Ficheux, D., Roques, B. P., Darlix, J. L., and Mely, Y. (2002). HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. *J Mol.Biol.* **317**, 385-399.

Betts, M. R., Krowka, J. F., Kepler, T. B., Davidian, M., Christopherson, C., Kwok, S., Louie, L., Eron, J., Sheppard, H., and Frelinger, J. A. (1999). Human immunodeficiency virus type 1-specific cytotoxic T lymphocyte activity is inversely correlated with HIV type 1 viral load in HIV type 1-infected long-term survivors. *Aids Res Hum Retrovirus* **15**, 1219-1228.

BHIVA (2001). British HIV Association (BHIVA) guidelines for the treatment of HIV- infected adults with antiretroviral therapy. *HIV.Med.* **2**, 276-313.

- Bieniasz, P. D., Fridell, R. A., Aramori, I., Ferguson, S. S. G., Caron, M. G., and Cullen, B. R. (1997). HIV-1-induced cell fusion is mediated by multiple regions within both the viral envelope and the CCR-5 co-receptor. *EMBO J* **16**, 2599-2609.
- Bjorndal, A., Deng, H. K., Jansson, M., Fiore, J. R., Colognesi, C., Karlsson, A., Albert, J., Scarlatti, G., Littman, D. R., and Fenyo, E. M. (1997). Coreceptor usage of primary human immunodeficiency virus type 1 isolates varies according to biological phenotype. *J. Virol.* **71**, 7478-7487.
- Blank, A., Gallant, J. A., Burgess, R. R., and Loeb, L. A. (1986). An RNA polymerase mutant with reduced accuracy of chain elongation. *Biochemistry* **25**, 5920-5928.
- Bobkov, A., Kazennova, E., Selimova, L., Bobkova, M., Khanina, T., Ladnaya, N., Kravchenko, A., Pokrovsky, V., Cheingsongpopov, R., and Weber, J. (1998). A sudden epidemic of HIV type 1 among injecting drug users in the former Soviet Union: Identification of subtype A, subtype B, and novel gagA/envB recombinants. *AIDS Res Hum Retroviruses* **14**, 669-676.
- Borman, A. M., Quillent, C., Charneau, P., Dauguet, C., and Clavel, F. (1995). Human immunodeficiency virus type 1 vif- mutant particles from restrictive cells: role of vif in correct particle assembly and infectivity. *J. Virol.* **69**, 2058-2067.
- Borrow, P., Lewicki, H., Wei, X. P., Horwitz, M. S., Peffer, N., Meyers, H., Nelson, J. A., Gairin, J. E., Hahn, B. H., Oldstone, M. B. A., and Shaw, G. M. (1997). Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nature Med* **3**, 205-211.
- Bowerman, B., Brown, P. O., Bishop, J. M., and Varmus, H. E. (1989). A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev.* **3**, 469-478.
- Boyd, M. T., Simpson, G. R., Cann, A. J., Johnson, M. A., and Weiss, R. A. (1993). A single amino acid substitution in the V1 loop of human immunodeficiency virus type 1 gp120 alters cellular tropism. *J. Virol.* **67**, 3649-3652.
- Bracho, M. A., Moya, A., and Barrio, E. (1998). Contribution of Taq polymerase-induced errors to the estimation of RNA virus diversity. *J. Gen. Virol.* **79**, 2921-2928.
- Brack-Werner, R. (1999). Astrocytes: HIV cellular reservoirs and important participants in neuropathogenesis. *AIDS* **13**, 1-22.
- Brelot, A., Heveker, N., Adema, K., Hosie, M. J., Willett, B., and Alizon, M. (1999). Effect of mutations in the second extracellular loop of CXCR4 on its utilization by human and feline immunodeficiency viruses. *J. Virol* **73**, 2576-2586.
- Broder, C. C., Dimitrov, D. S., Blumenthal, R., and Berger, E. A. (1993). The block to HIV-1 envelope glycoprotein-mediated membrane fusion in animal cells expressing human CD4 can be overcome by a human cell component (s). *Virology* **193**, 483-491.
- Bron, R., Klasse, P. J., Wilkinson, D., Clapham, P. R., Pelchenmatthews, A., Power, C., Wells, T. N. C., Kim, J., Peiper, S. C., Hoxie, J. A., and Marsh, M. (1997). Promiscuous use of CC and CXC chemokine receptors in cell- to- cell fusion mediated by a human immunodeficiency virus type 2 envelope protein. *J. Virol.* **71**, 8405-8415.
- Brown, A. J. (1997). Analysis of HIV-1 env gene sequences reveals evidence for a low effective number in the viral population. *Proc. Natl. Acad. Sci. U.S.A* **94**, 1862-1865.
- Brown, A. J. L. and Cleland, A. (1996). Independent evolution of the env and pol genes of HIV-1 during zidovudine therapy. *AIDS* **10**, 1067-1073.

- Brown, A. J. L. and Richman, D. D. (1997). HIV-1: Gambling on the evolution of drug resistance? *Nature Med* **3**, 268-271.
- Buiser, R. G., Bambara, R. A., and Fay, P. J. (1993). Pausing by retroviral DNA polymerases promotes strand transfer from internal regions of RNA donor templates to homopolymeric acceptor templates. *Biochim.Biophys.Acta* **1216**, 20-30.
- Bukrinsky, M. I., Haggerty, S., Dempsey, M. P., Sharova, N., Adzhubei, A., Spitz, L., Lewis, P., Goldfarb, D., Emerman, M., and Stevenson, M. (1993a). A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. *Nature* **365**, 666-669.
- Bukrinsky, M. I., Sharova, N., McDonald, T. L., Pushkarskaya, T., Tarpley, W. G., and Stevenson, M. (1993b). Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type-1 with viral nucleic acids following acute infection. *Proc.Natl.Acad.Sci.USA* **90**, 6125-6129.
- Bullough, P. A., Hughson, F. M., Skehel, J. J., and Wiley, D. C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**, 37-43.
- Burns, D. K., Risser, R. C., and White, C. L., III (1991). The neuropathology of human immunodeficiency virus infection. The Dallas, Texas, experience. *Arch.Pathol.Lab Med.* **115**, 1112-1124.
- Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P., and Craigie, R. (1993). Domains of the integrase protein of human immunodeficiency virus type-1 responsible for polynucleotidyl transfer and zinc binding. *Proc.Natl.Acad.Sci.USA* **90**, 3428-3432.
- Canki, M., Thai, J. N., Chao, W., Ghorpade, A., Potash, M. J., and Volsky, D. J. (2001). Highly productive infection with pseudotyped human immunodeficiency virus type 1 (hiv-1) indicates no intracellular restrictions to hiv-1 replication in primary human astrocytes. *J.Virol.* **75**, 7925-33.
- Cann, A. J., Churcher, M. J., Boyd, M., O'Brien, W., Zhao, J. Q., Zack, J., and Chen, I. S. (1992). The region of the envelope gene of human immunodeficiency virus type 1 responsible for determination of cell tropism. *J.Virol.* **66**, 305-309.
- Cantin, R., Fortin, J. F., Lamontagne, G., and Tremblay, M. (1997). The presence of host-derived HLA-DR1 on human immunodeficiency virus type 1 increases viral infectivity. *J Virol* **71**, 1922-1930.
- Cao, J., Sullivan, N., Desjardin, E., Parolin, C., Robinson, J., Wyatt, R., and Sodroski, J. (1997). Replication and neutralization of human immunodeficiency virus type 1 lacking the V1 and V2 variable loops of the gp120 envelope glycoprotein. *J Virol* **71**, 9808-9812.
- Cao, Y. Z., Qin, L. M., Zhang, L. Q., Safrit, J., and Ho, D. D. (1995). Virologic and immunologic characterization of long-term survivors of human immunodeficiency virus type 1 infection. *N.Engl.J.Med.* **332**, 201-208.
- Carr, J. K., Salminen, M. O., Albert, J., Sandersbuell, E., Gotte, D., Birx, D. L., and McCutchan, F. E. (1998). Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants. *Virology* **247**, 22-31.
- Carr, J. K., Salminen, M. O., Koch, C., Gotte, D., Artenstein, A. W., Hegerich, P. A., StLouis, D., Burke, D. S., and McCutchan, F. E. (1996). Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J.Virol.* **70**, 5935-5943.

- Carrillo, A. and Ratner, L. (1996). Cooperative effects of the human immunodeficiency virus type 1 envelope variable loops V1 and V3 in mediating infectivity for T cells. *Journal of Virology* **70**, 1310-1316.
- Casella, C. R., Raffini, L. J., and Panganiban, A. T. (1997). Pleiotropic mutations in the HIV-1 matrix protein that affect diverse steps in replication. *Virology* **228**, 294-306.
- CDC (1981). *Pneumocystis pneumonia*-Los Angeles. *Morbidity and Mortality Weekly Report* **30**, 250-252.
- Cepko, C. (2001). Introduction of DNA into Mammalian Cells. In "Current Protocols in Molecular Biology" (F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl, Eds.), p. 9.0.1-9.14.6. John Wiley and Sons, Inc., USA.
- Chan, D. C., Chutkowski, C. T., and Kim, P. S. (1998). Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. *Proc.Natl.Acad.Sci.U.S.A* **95**, 15613-15617.
- Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997). Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**, 263-273.
- Chan, S. Y., Speck, R. F., Power, C., Gaffen, S. L., Chesebro, B., and Goldsmith, M. A. (1999). V3 recombinants indicate a central role for CCR5 as a coreceptor in tissue infection by human immunodeficiency virus type 1. *J Virol* **73**, 2350-2358.
- Chang, J., Jozwiak, R., Wang, B., Ng, T., Ge, Y. C., Bolton, W., Dwyer, D. E., Randle, C., Osborn, R., Cunningham, A. L., and Saksena, N. K. (1998). Unique HIV type 1 V3 region sequences derived from six different regions of brain: Region-specific evolution within host- determined quasispecies - Short communication. *AIDS Res Hum Retroviruses* **14**, 25-30.
- Charneau, P., Alizon, M., and Clavel, F. (1992). A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication. *J.Virol.* **66**, 2814-2820.
- Charneau, P., Mirambeau, G., Roux, P., Paulous, S., Buc, H., and Clavel, F. (1994). HIV-1 reverse transcription. A termination step at the center of the genome. *J.Mol.Biol.* **241**, 651-662.
- Chen, H., Wood, C., and Petit, C. K. (2000). Comparisons of HIV-1 viral sequences in brain, choroid plexus and spleen: Potential role of choroid plexus in the pathogenesis of HIV encephalitis. *J.Neurovirol.* **6**, 498-506.
- Chen, M. Y., Maldarelli, F., Karczewski, M. K., Willey, R. L., and Strebel, K. (1993). Human immunodeficiency virus type-1 vpu protein induces degradation of CD4 invitro - the cytoplasmic domain of CD4 contributes to vpu sensitivity. *J.Virol.* **67**, 3877-3884.
- Chen, Z. W., Kwon, D., Jin, Z. Q., Monard, S., Telfer, P., Jones, M. S., Lu, C. Y., Aguilar, R. F., Ho, D. D., and Marx, P. A. (1998). Natural infection of a homozygous Delta 24 CCR5 red-capped mangabey with an R2b-tropic simian immunodeficiency virus. *J Exp Med* **188**, 2057-2065.
- Cheng-Mayer, C. and Levy, J. A. (1988). Distinct biologic and serologic properties of HIV isolates from the brain. *Ann.Neurol.* **23**, s58-s61.
- Cheng Mayer, C., Quiroga, M., Tung, J. W., Dina, D., and Levy, J. A. (1990). Viral determinants of human immunodeficiency virus type 1 T-cell or macrophage tropism, cytopathogenicity, and CD4 antigen modulation. *J.Virol.* **64**, 4390-4398.
- Cheng Mayer, C., Rutka, J. T., Rosenblum, M. L., McHugh, T., Stites, D. P., and Levy, J. A. (1987). The Human Immunodeficiency Virus can productively infect cultured human glial cells. *Proc.Natl.Acad.Sci.USA* **84**, 3526-3530.

- Chesebro, B., Buller, R., Portis, J., and Wehrly, K. (1990). Failure of human immunodeficiency virus entry and infection in CD4-positive human brain and skin cells. *J. Virol.* **64**, 215-221.
- Chesebro, B., Nishio, J., Perryman, S., Cann, A., O'Brien, W., Chen, I. S., and Wehrly, K. (1991). Identification of human immunodeficiency virus envelope gene sequences influencing viral entry into CD4-positive Hela cells, T-leukemia cells and macrophages. *J. Virol.* **65**, 5782-5789.
- Chesebro, B., Wehrly, K., Nishio, J., and Perryman, S. (1992). 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.
- Chiodi, F., Albert, J., Olausson, E., Norkrans, G., Hagberg, L., Sonnerborg, A., Asjo, B., and Fenyo, E. M. (1988). Isolation frequency of human immunodeficiency virus from cerebrospinal fluid and blood of patients with varying severity of HIV infection. *AIDS Res. Hum. Retroviruses* **4**, 351-358.
- Chiodi, F., Fuerstenberg, S., Gidlund, M., Asjo, B., and Fenyo, E. M. (1987). Infection of brain-derived cells with the human immunodeficiency virus. *J. Virol.* **61**, 1244-1247.
- Chiodi, F., Keys, B., Albert, J., Hagberg, L., Lundeberg, J., Uhlen, M., Fenyo, E. M., and Norkrans, G. (1992). Human immunodeficiency virus type 1 is present in the cerebrospinal fluid of a majority of infected individuals. *Journal of Clinical Microbiology* **30**, 1768-1771.
- Cho, M. W., Lee, M. K., Carney, M. C., Berson, J. F., Doms, R. W., and Martin, M. A. (1998). Identification of determinants on a dualtropic human immunodeficiency virus type 1 envelope glycoprotein that confer usage of CXCR4. *J. Virol.* **72**, 2509-2515.
- Choe, H., Farzan, M., Konkel, M., Martin, K., Sun, Y., Marcon, L., Cayabyab, M., Berman, M., Dorf, M. E., Gerard, N., Gerard, C., and Sodroski, J. (1998). The orphan seven-transmembrane receptor Apj supports the entry of primary T-cell-line-tropic and dualtropic human immunodeficiency virus type 1. *J. Virol.* **72**, 6113-6118.
- Choe, H., Farzan, M., Sun, Y., Sullivan, N., Rollins, B., Ponath, P. D., Wu, L. J., Mackay, C. R., Larosa, G., Newman, W., Gerard, N., Gerard, C., and Sodroski, J. (1996). The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell* **85**, 1135-1148.
- Choe, W., Albright, A., Sulcove, J., Jaffer, S., Hesselgesser, J., Lavi, E., Crino, P., and Kolson, D. L. (2000). Functional expression of the seven-transmembrane HIV-1 co-receptor APJ in neural cells. *J. Neurovirol.* **6 Suppl 1:S61-9**, S61-9.
- Chowers, M. Y., Pandori, M. W., Spina, C. A., Richman, D. D., and Guatelli, J. C. (1995). The growth advantage conferred by HIV-1 nef is determined at the level of viral DNA formation and is independent of CD4 downregulation. *Virology* **212**, 451-457.
- Chowers, M. Y., Spina, C. A., Kwok, T. J., Fitch, N. J. S., Richman, D. D., and Guatelli, J. C. (1994). Optimal infectivity in vitro of human immunodeficiency virus type 1 requires an intact nef gene. *J. Virol.* **68**, 2906-2914.
- Clapham, P. R., Blanc, D., and Weiss, R. A. (1991). Specific cell surface requirements for the infection of CD4-positive cells by human immunodeficiency virus types 1 and 2 and by Simian immunodeficiency virus. *Virology* **181**, 703-715.
- Clapham, P. R. and McKnight, A. (2002). Cell surface receptors, virus entry and tropism of primate lentiviruses. *J. Gen. Virol.* **83**, 1809-1829.

- Clark, S. J., Saag, M. S., Decker, W. D., Campbell-Hill, S., Robertson, J. L., Veldkamp, P. J., Kappes, J. C., Hahn, B. H., and Shaw, G. M. (1991). High titers of cytopathic virus in plasma of patients with symptomatic primary HIV-1 infection. *N.Engl.J.Med.* **324**, 954-960.
- Clavel, F., Guetard, D., Brun Vezinet, F., Chamaret, S., Rey, M. A., Santos Ferreira, M. O., Laurent, A. G., Dauguet, C., Katlama, C., Rouzioux, C., and et al (1986a). Isolation of a new human retrovirus from West African patients with AIDS. *Science.* **233**, 343-346.
- Clavel, F., Guyader, M., Guetard, D., Salle, M., Montagnier, L., and Alizon, M. (1986b). Molecular cloning and polymorphism of the human immune deficiency virus type 2. *Nature.* **324**, 691-695.
- Clavel, F., Hoggan, M. D., Willey, R. L., Strebel, K., Martin, M. A., and Repaske, R. (1989). Genetic Recombination of Human Immunodeficiency Virus. *Journal of Virology* **63**, 1455-1459.
- Clever, J. L. and Parslow, T. G. (1997). Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J Virol* **71**, 3407-3414.
- Coates, T. J., Aggleton, P., Gutzwiller, F., DesJarlais, D., Kihara, M., Kippax, S., Schechter, M., and Vandenhoeck, J. A. R. (1996). HIV prevention in developed countries. *Lancet* **348**, 1143-1148.
- Coffin, J., Haase, A., Levy, J. A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., Vogt, P., and et al (1986). What to call the AIDS virus? *Nature.* **321**, 10.
- Coffin, J. M. (1979). Review: Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *J.Gen.Virol.* **42**, 1-26.
- Coffin, J. M. (1996). Retroviridae: The viruses and their replication. In "Fields Virology" (B. N. Fields, D. M. Knipe, P. M. Howley, R. M. Chanock, J. L. Melnick, T. P. Monath, B. Roizman, and S. E. Straus, Eds.), Vol. 2, pp. 1767-1848. Lippincott-Raven, Philadelphia.
- Coffin, J. M. (1992). Genetic diversity and evolution of retroviruses. *Curr.Top.Microbiol.Immunol.* **176**, 143-164.
- Coffin, J. M. (1995). HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science* **267**, 483-489.
- Cohen, E. A., Dehni, G., Sodroski, J. G., and Haseltine, W. A. (1990). Human immunodeficiency virus vpr product is a virion-associated regulatory protein. *J.Virol.* **64**, 3097-3099.
- Cohen, E. A., Terwilliger, E. F., Sodroski, J. G., and Haseltine, W. A. (1988). Identification of a protein encoded by the vpu gene of HIV-1. *Nature* **334**, 532-534.
- Cohen, O. J. and Fauci, A. S. (2001). Pathogenesis and Medical Aspects of HIV-1 Infection. In "Fields Virology" (D. M. Knipe and P. M. Howley, Eds.), pp. 2043-2094. Lippincott Williams and Wilkins, Philadelphia.
- Collins, K. L., Chen, B. K., Kalams, S. A., Walker, B. D., and Baltimore, D. (1998). HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature* **391**, 397-401.
- Copeland, T. D. and Oroszlan, S. (1988). Genetic locus, primary structure, and chemical synthesis of human immunodeficiency virus protease. *Gene Anal.Tech.* **5**, 109-115.
- Corbet, S., MullerTrutwin, M. C., Versmisse, P., Delarue, S., Ayouba, A., Lewis, J., Brunak, S., Martin, P., Brunvezinet, F., Simon, F., Barresinoussi, F., and Mauciere, P. (2000). env Sequences of simian immunodeficiency viruses from chimpanzees in Cameroon are strongly related to those of human immunodeficiency virus group N from the same geographic area. *J Virol* **74**, 529-534.

- Cowan, M. J., Hellmann, D., Chudwin, D., Wara, D. W., Chang, R. S., and Ammann, A. J. (1984). Maternal transmission of acquired immune deficiency syndrome. *Pediatrics* **73**, 382-386.
- Craigie, R. (1992). Hotspots and warm spots: integration specificity of retroelements. *Trends Genet.* **8**, 187-190.
- Cruikshank, W. W., Kornfeld, H., and Center, D. M. (1998). Signaling and functional properties of interleukin-16. *Int.Rev.Immunol.* **16**, 523-540.
- Cunningham, T. P., Montelaro, R. C., and Rushlow, K. E. (1993). Lentivirus envelope sequences and proviral genomes are stabilized in escherichia-coli when cloned in low-copy- number plasmid vectors. *Gene* **124**, 93-98.
- Curtis, B. M., Scharnowske, S., and Watson, A. J. (1992). Sequence and expression of a membrane-associated C-type lectin that exhibits CD4-independent binding of human immunodeficiency virus envelope glycoprotein gp120. *Proc.Natl.Acad.Sci.U.S.A* **89**, 8356-8360.
- D'Souza, M. P., Cairns, J. S., and Plaeger, S. F. (2000). Current evidence and future directions for targeting HIV entry: therapeutic and prophylactic strategies. *JAMA* **284**, 215-222.
- Daar, E. S., Moudgil, T., Meyer, R. D., and Ho, D. D. (1991). Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection. *N.Engl.J.Med.* **324**, 961-964.
- Dalglish, A. G., Beverley, P. C., Clapham, P. R., Crawford, D. H., Greaves, M. F., and Weiss, R. A. (1984). The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature.* **312**, 763-767.
- Dallasta, L. M., Pizarov, L. A., Esplen, J. E., Werley, J. V., Moses, A. V., Nelson, J. A., and Achim, C. L. (1999). Blood-brain barrier tight junction disruption in human immunodeficiency virus-1 encephalitis. *Am.J.Pathol.* **155**, 1915-1927.
- Danieli, T., Pelletier, S. L., Henis, Y. I., and White, J. M. (1996). Membrane fusion mediated by the influenza virus hemagglutinin requires the concerted action of at least three hemagglutinin trimers. *J.Cell Biol.* **133**, 559-569.
- Davis, C. B., Dikic, I., Unutmaz, D., Hill, C. M., Arthos, J., Siani, M. A., Thompson, D. A., Schlessinger, J., and Littman, D. R. (1997). Signal transduction due to HIV-1 envelope interactions with chemokine receptors CXCR4 or CCR5. *J Exp Med* **186**, 1793-1798.
- Davis, L. E., Hjelle, B. L., and Miller, V. E. (1992). Early brain invasion in iatrogenic human immunodeficiency virus infection. *Neurology* **42**, 1736-1739.
- de Jong, J. J., de Ronde, A., Keulen, W., Tersmette, M., and Goudsmit, J. (1992a). Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. *J.Virol.* **66** , 6777-6780.
- de Jong, J. J., Goudsmit, J., Keulen, W., Klaver, B., Krone, W., and (1992b). Human immunodeficiency virus type 1 clones chimeric for the envelope V3 domain differ in syncytium formation and replication capacity. *J.Virol.* **66**, 757-765.
- Deacon, N. J., Tsykin, A., Solomon, A., Smith, K., Ludfordmenting, M., Hooker, D. J., McPhee, D. A., Greenway, A. L., Ellett, A., Chatfield, C., Lawson, V. A., Crowe, S., Maerz, A., Sonza, S., Learmont, J., Sullivan, J. S., Cunningham, A., Dwyer, D., Downton, D., and Mills, J. (1995). Genomic structure of an attenuated quasi species of HIV-1 from a blood transfusion donor and recipients. *Science* **270**, 988-991.

- Dean, M., Carrington, M., Winkler, C., Huttley, G. A., Smith, M. W., Allikmets, R., Goedert, J. J., Buchbinder, S. P., Vittinghoff, E., Gomperts, E., Donfield, S., Vlahov, D., Kaslow, R., Saah, A., Rinaldo, C., Detels, R., and Obrien, S. J. (1996). Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. *Science* **273**, 1856-1862.
- Debouck, C., Gorniak, J. G., Strickler, J. E., Meek, T. D., Metcalf, B. W., and Rosenberg, M. (1987). Human immunodeficiency virus protease expressed in *Escherichia coli* exhibits autoprocessing and specific maturation of the gag precursor. *Proc.Natl.Acad.Sci.U.S.A* **84**, 8903-8906.
- DeGuzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998). Structure of the HIV-1 nucleocapsid protein bound to the SL3 Psi- RNA recognition element. *Science* **279**, 384-388.
- Deleys, R., Vanderborght, B., Vandenaeseveld, M., Heyndrickx, L., Van Geel, A., Wauters, C., Bernaerts, R., Saman, E., Nijs, P., Willems, B., Taelman, H., van der Groen, G., Piot, P., Tersmette, T., Huisman, J. G., and Van Heuverswyn, H. (1990). Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of West-Central African origin. *J.Virol.* **64**, 1207-1216.
- Delezay, O., Koch, N., Yahi, N., Hammache, D., Tourres, C., Tamalet, C., and Fantini, J. (1997). Co-expression of CXCR4/fusin and galactosylceramide in the human intestinal epithelial cell line HT-29. *AIDS* **11**, 1311-1318.
- Delviks, K. A. and Pathak, V. K. (1999). Effect of distance between homologous sequences and 3' homology on the frequency of retroviral reverse transcriptase template switching. *J.Virol.* **73**, 7923-7932.
- Delwart, E. L., Mullins, J. I., Gupta, P., Learn, G. H., Holodniy, M., Katzenstein, D., Walker, B. D., and Singh, M. K. (1998). Human immunodeficiency virus type 1 populations in blood and semen. *J Virol* **72**, 617-623.
- Delwart, E. L., Pan, H., Sheppard, H. W., Wolpert, D., Neumann, A. U., Korber, B., and Mullins, J. I. (1997). Slower evolution of human immunodeficiency virus type I quasispecies during progression to AIDS. *J.Virol.* **71**, 7498-7508.
- Deng, H. K., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Dimarzio, P., Marmon, S., Sutton, R. E., Hill, C. M., Davis, C. B., Peiper, S. C., Schall, T. J., Littman, D. R., and Landau, N. R. (1996). Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**, 661-666.
- Deng, H. K., Unutmaz, D., Kewalramani, V. N., and Littman, D. R. (1997). Expression cloning of new receptors used by simian and human immunodeficiency viruses. *Nature* **388**, 296-300.
- Desrosiers, R. C. and Letvin, N. L. (1987). Animal models for acquired immunodeficiency syndrome. *Rev.Infect.Dis.* **9**, 438-446.
- Destefano, J. J., Buiser, R. G., Mallaber, L. M., Fay, P. J., and Bambara, R. A. (1992a). Parameters that influence processive synthesis and site-specific termination by human immunodeficiency virus reverse transcriptase on RNA and DNA templates. *Biochim.Biophys.Acta* **1131**, 270-280.
- Destefano, J. J., Mallaber, L. M., Rodriguez-Rodriguez, L., Fay, P. J., and Bambara, R. A. (1992b). Requirements for strand transfer between internal regions of heteropolymer templates by human immunodeficiency virus reverse transcriptase. *J.Virol.* **66**, 6370-6378.
- Di Stefano, M., Wilt, S., Gray, F., DuboisDalcq, M., and Chiodi, F. (1996). HIV type 1 V3 sequences and the development of dementia during AIDS. *AIDS Res.Hum.Retroviruses* **12**, 471-476.

- Diaz, L. and Destefano, J. J. (1996). Strand transfer is enhanced by mismatched nucleotides at the 3' primer terminus: A possible link between HIV reverse transcriptase fidelity and recombination. *Nucleic Acids Research* **24**, 3086-3092.
- Diaz, R. S., Sabino, E. C., Mayer, A., Mosley, J. W., and Busch, M. P. (1995). Dual human immunodeficiency virus type 1 infection and recombination in a dually exposed transfusion recipient. *J.Virol.* **69**, 3273-3281.
- Dick, A. D., Pell, M., Brew, B. J., Foulcher, E., and Sedgwick, J. D. (1997). Direct ex vivo flow cytometric analysis of human microglial cell CD4 expression: examination of central nervous system biopsy specimens from HIV-seropositive patients and patients with other neurological disease. *AIDS* **11**, 1699-1708.
- Dimitrov, D. S., Norwood, D., Stantchev, T. S., Feng, Y. R., Xiao, X. D., and Broder, C. C. (1999). A mechanism of resistance to HIV-1 entry: Inefficient interactions of CXCR4 with CD4 and gp120 in macrophages. *Virology* **259**, 1-6.
- Dimitrov, D. S., Xiao, X., Chabot, D. J., and Broder, C. C. (1998). HIV coreceptors. *J.Membr.Biol.* **166**, 75-90.
- Dingwall, C., Ernberg, I., Gait, M. J., Green, S. M., Heaphy, S., Karn, J., Lowe, A. D., Singh, M., Skinner, M. A., and Valerio, R. (1989). Human immunodeficiency virus 1 tat protein binds trans-activation- responsive region (TAR) RNA in vitro. *Proc.Natl.Acad.Sci.U.S.A* **86**, 6925-6929.
- Dittmar, M. T., McKnight, A., Simmons, G., Clapham, P. R., Weiss, R. A., and Simmonds, P. (1997). HIV-1 tropism and co-receptor use. *Nature* **385**, 495-496.
- Domingo, E. and Holland, J. J. (1997). RNA virus mutations and fitness for survival. *Annu.Rev.Microbiol.* **51**, 151-178.
- Doms, R. W. (2000). Beyond receptor expression: the influence of receptor conformation, density, and affinity in HIV-1 infection. *Virology* **276**, 229-237.
- Doms, R. W. and Moore, J. P. (2000). HIV-1 membrane fusion: targets of opportunity. *J.Cell Biol.* **151**, F9-14.
- Doms, R. W. and Peipert, S. C. (1997). Unwelcomed guests with master keys: How HIV uses chemokine receptors for cellular entry. *Virology* **235**, 179-190.
- Donaldson, Y. K., Bell, J. E., Holmes, E. C., Hughes, E. S., Brown, H. K., and Simmonds, P. (1994a). In vivo distribution and cytopathology of variants of human immunodeficiency virus type 1 showing restricted sequence variability in the V3 loop. *J.Virol.* **68** , 5991-6005.
- Donaldson, Y. K., Bell, J. E., Ironside, J. W., Brettle, R. P., Robertson, J. R., Busuttill, A., and Simmonds, P. (1994b). Redistribution of HIV outside the lymphoid system with onset of AIDS. *Lancet* **343**, 382-385.
- Doranz, B. J., Rucker, J., Yi, Y. J., Smyth, R. J., Samson, M., Peiper, S. C., Parmentier, M., Collman, R. G., and Doms, R. W. (1996). 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**, 1149-1158.
- Dore, G. J., Correll, P. K., Li, Y. M., Kaldor, J. M., Cooper, D. A., and Brew, B. J. (1999). Changes to AIDS dementia complex in the era of highly active antiretroviral therapy. *AIDS* **13**, 1249-1253.
- Dougherty, J. P. and Temin, H. M. (1986). High mutation rate of a spleen necrosis virus-based retrovirus vector. *Mol.Cell Biol.* **6**, 4387-4395.

- Dragic, T., Charneau, P., Clavel, F., and Alizon, M. (1992). Complementation of murine cells for human immunodeficiency virus envelope/CD4-mediated fusion in human/murine heterokaryons. *J. Virol.* **66**, 4794-4802.
- Dragic, T., Litwin, V., Allaway, G. P., Martin, S. R., Huang, Y. X., Nagashima, K. A., Cayanan, C., Maddon, P. J., Koup, R. A., Moore, J. P., and Paxton, W. A. (1996). HIV-1 entry into CD4 (+) cells is mediated by the chemokine receptor CC-CKR-5. *Nature* **381**, 667-673.
- DSouza, M. P. and Harden, V. A. (1996). Chemokines and HIV-1 second receptors - Confluence of two fields generates optimism in AIDS research. *Nature Med* **2**, 1293-1300.
- Dumonceaux, J., Nisole, S., Chanel, C., Quivet, L., Amara, A., Baleux, F., Briand, P., and Hazan, U. (1998). Spontaneous mutations in the env gene of the human immunodeficiency virus type 1 NDK isolate are associated with a CD4-independent entry phenotype. *J Virol* **72**, 512-519.
- Duong, T., Ades, A. E., Gibb, D. M., Tookey, P. A., and Masters, J. (1999). Vertical transmission rates for HIV in the British Isles: estimates based on surveillance data. *BMJ* **319**, 1227-1229.
- Dutch, R. E., Jardetzky, T. S., and Lamb, R. A. (2000). Virus membrane fusion proteins: biological machines that undergo a metamorphosis. *Biosci.Rep.* **20**, 597-612.
- Dykes, C., Mootsikapun, P., Dexter, A., Berrios, L., Chiulli, M., Reichman, R. C., and Demeter, L. M. (2000). Analysis of env sequence evolution in human immunodeficiency virus- infected patients receiving therapy with nonnucleoside reverse- transcriptase inhibitors. *Journal of Infectious Diseases* **182**, 316-320.
- Earl, P. L., Doms, R. W., and Moss, B. (1990). Oligomeric structure of the human immunodeficiency virus type 1 envelope glycoprotein. *Proc.Natl.Acad.Sci.U.S.A.* **87**, 648-652.
- Earl, P. L., Moss, B., and Doms, R. W. (1991). Folding, interaction with GRP78-BiP, assembly, and transport of the human immunodeficiency virus type 1 envelope protein. *J. Virol.* **65**, 2047-2055.
- Edinger, A. L., Hoffman, T. L., Sharron, M., Lee, B. H., Yi, Y. J., Choe, W., Kolson, D. L., Mitrovic, B., Zhou, Y. Q., Faulds, D., Collman, R. G., Hesselgesser, J., Horuk, R., and Doms, R. W. (1998). An orphan seven-transmembrane domain receptor expressed widely in the brain functions as a coreceptor for human immunodeficiency virus type 1 and simian immunodeficiency virus. *J Virol* **72**, 7934-7940.
- Edinger, A. L., Mankowski, J. L., Doranz, B. J., Margulies, B. J., Lee, B., Rucker, J., Sharron, M., Hoffman, T. L., Berson, J. F., Zink, M. C., Hirsch, V. M., Clements, J. E., and Doms, R. W. (1997). CD4-independent, CCR5-dependent infection of brain capillary endothelial cells by a neurovirulent simian immunodeficiency virus strain. *Proc Natl Acad Sci USA* **94**, 14742-14747.
- Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465-523.
- Eigen, M. (1996). On the nature of virus quasispecies. *Trends Microbiol.* **4**, 216-218.
- Eigen, M. and Schuster, P. (1977). The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften* **64**, 541-565.
- Embreton, J., Zupancic, M., Beneke, J., Till, M., Wolinsky, S., Ribas, J. L., Burke, A., and Haase, A. T. (1993). Analysis of human immunodeficiency virus-infected tissues by amplification and insitu hybridization reveals latent and permissive infections at single-cell resolution. *Proc.Natl.Acad.Sci.USA* **90**, 357-361.

- Engelman, A., Hickman, A. B., and Craigie, R. (1994). The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. *J. Virol.* **68**, 5911-5917.
- Enting, R. H., Hoetelmans, R. M. W., Lange, J. M. A., Burger, D. M., Beijnen, J. H., and Portegies, P. (1998). Antiretroviral drugs and the central nervous system. *AIDS* **12**, 1941-1955.
- Epstein, L. G. and Gendelman, H. E. (1993). Human immunodeficiency virus type 1 infection of the nervous system: pathogenetic mechanisms. *Ann. Neurol.* **33**, 429-436.
- Epstein, L. G., Kuiken, C., Blumberg, B. M., Hartman, S., Sharer, L. R., Clement, M., and Goudsmit, J. (1991a). HIV-1 V3 domain variation in brain and spleen of children with AIDS: tissue-specific evolution within host-determined quasispecies. *Virology.* **180**, 583-590.
- Epstein, L. G., Kuiken, C., Blumberg, B. M., Hartman, S., Sharer, L. R., Clement, M., and Goudsmit, J. (1991b). HIV-1 V3 domain variation in brain and spleen of children with AIDS: tissue-specific evolution within host-determined quasispecies. *Virology* **180**, 583-590.
- Erie, D. A., Hajiseyedjavadi, O., Young, M. C., and von Hippel, P. H. (1993). Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription. *Science* **262**, 867-873.
- Ewart, G. D., Sutherland, T., Gage, P. W., and Cox, G. B. (1996). The Vpu protein of human immunodeficiency virus type 1 forms cation-selective ion channels. *J. Virol.* **70**, 7108-7115.
- Falangola, M. F., Hanly, A., Galvao-Castro, B., and Petito, C. K. (1995). HIV infection of human choroid plexus: a possible mechanism of viral entry into the CNS. *J. Neuropathol. Exp. Neurol.* **54**, 497-503.
- Fang, G., Zhu, G., Burger, H., Keithly, J. S., and Weiser, B. (1998). Minimizing DNA recombination during long RT-PCR. *J. Virol. Methods* **76**, 139-148.
- Fantini, J., Cook, D. G., Nathanson, N., Spitalnik, S. L., and Gonzalezscarano, F. (1993). Infection of colonic epithelial cell lines by type-1 human immunodeficiency virus is associated with cell surface expression of galactosylceramide, a potential alternative gp120 receptor. *Proc. Natl. Acad. Sci. USA* **90**, 2700-2704.
- Farzan, M., Choe, H., Martin, K., Marcon, L., Hofmann, W., Karlsson, G., Sun, Y., Barrett, P., Marchand, N., Sullivan, N., Gerard, N., Gerard, C., and Sodroski, J. (1997). Two orphan seven-transmembrane segment receptors which are expressed in CD4-positive cells support simian immunodeficiency virus infection. *J. Exp. Med* **186**, 405-411.
- Farzan, M., Choe, H., Vaca, L., Martin, K., Sun, Y., Desjardins, E., Ruffing, N., Wu, L. J., Wyatt, R., Gerard, N., Gerard, C., and Sodroski, J. (1998). A tyrosine-rich region in the N terminus of CCR5 is important for human immunodeficiency virus type 1 entry and mediates an association between gp120 and CCR5. *J. Virol* **72**, 1160-1164.
- Fauci, A. S. and Desrosiers, R. C. (1997). Pathogenesis of HIV and SIV. In "Retroviruses" (J. M. Coffin, S. H. Hughes, and H. E. Varmus, Eds.), pp. 587-636. Cold Spring Harbor Laboratory Press, New York.
- Fauci, A. S. (1988). The human immunodeficiency virus: infectivity and mechanisms of pathogenesis. *Science.* **239**, 617-622.
- Feinberg, M. B. (1996). Changing the natural history of HIV disease. *Lancet* **348**, 239-246.

- Feinberg, M. B., Baltimore, D., and Frankel, A. D. (1991). The role of Tat in the human immunodeficiency virus life cycle indicates a primary effect on transcriptional elongation. *Proc.Natl.Acad.Sci.U.S.A* **88**, 4045-4049.
- Feinstein, S. C., Ross, S. R., and Yamamoto, K. R. (1982). Chromosomal position effects determine transcriptional potential of integrated mammary tumor virus DNA. *J.Mol.Biol.* **156** , 549-565.
- Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**, 783-791.
- Felsenstein, J. (1988). Phylogenies from molecular sequences: inference and reliability. *Annu.Rev.Genet.* **22**, 521-565.
- Felsenstein, J. (1989). PHYLIP-phylogeny inference package (version 3.2). *Cladistics* **5**, 164-166.
- Felzien, L. K., Woffendin, C., Hottiger, M. O., Subbramanian, R. A., Cohen, E. A., and Nabel, G. J. (1998). HIV transcriptional activation by the accessory protein, VPR, is mediated by the p300 co-activator. *Proc Natl Acad Sci USA* **95**, 5281-5286.
- Feng, S. and Holland, E. C. (1988). HIV-1 tat trans-activation requires the loop sequence within tar. *Nature* **334**, 165-167.
- Feng, Y., Broder, C. C., Kennedy, P. E., and Berger, E. A. (1996). HIV-1 entry cofactor: Functional cDNA cloning of a seven- transmembrane, G protein-coupled receptor. *Science* **272**, 872-877.
- Fenyo, E. M., Morfeldt Manson, L., Chiodi, F., Lind, B., von Gegerfelt, A., Albert, J., Olausson, E., and Asjo, B. (1988). Distinct replicative and cytopathic characteristics of human immunodeficiency virus isolates. *J.Virol.* **62**, 4414-4419.
- Ferrando, S., vanGorp, W., McElhiney, M., Goggin, K., Sewell, M., and Rabkin, J. (1998). Highly active antiretroviral treatment in HIV infection: benefits for neuropsychological function. *AIDS* **12**, F65-F70.
- Fischer, U., Huber, J., Boelens, W. C., Mattaj, I. W., and Luhrmann, R. (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.
- Fisher, A. G., Ensoli, B., Looney, D., Rose, A., Gallo, R. C., Saag, M. S., Shaw, G. M., Hahn, B. H., and Wong Staal, F. (1988). Biologically diverse molecular variants within a single HIV-1 isolate. *Nature* **334**, 444-447.
- Forns, X., Bukh, J., Purcell, R. H., and Emerson, S. U. (1997). How Escherichia coli can bias the results of molecular cloning: preferential selection of defective genomes of hepatitis C virus during the cloning procedure. *Proc.Natl.Acad.Sci.U.S.A* **94**, 13909-13914.
- Fortin, J. F., Barbeau, B., Hedman, H., Lundgren, E., and Tremblay, M. J. (1999). Role of the leukocyte function antigen-1 conformational state in the process of human immunodeficiency virus type 1-mediated syncytium formation and virus infection. *Virology* **257**, 228-238.
- Fouchier, R. A. M., Broersen, S. M., Brouwer, M., Tersmette, M., Vantwout, A. B., Groenink, M., and Schuitemaker, H. (1995a). Temporal relationship between elongation of the HIV type 1 glycoprotein 120 V2 domain and the conversion toward a syncytium- inducing phenotype. *AIDS Res.Hum.Retroviruses* **11**, 1473-1478.
- Fouchier, R. A. M., Brouwer, M., Broersen, S. M., and Schuitemaker, H. (1995b). Simple determination of human immunodeficiency virus type 1 syncytium-inducing v3 genotype by PCR. *Journal of Clinical Microbiology* **33**, 906-911.

- Fouchier, R. A. M., Groenink, M., Kootstra, N. A., Tersmette, M., Huisman, H. G., Miedema, F., and Schuitemaker, H. (1992). Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J. Virol* **66**, 3138-3187.
- Freed, E. O. and Martin, M. A. (1994). HIV-1 infection of non-dividing cells. *Nature* **369**, 107-108.
- Freed, E. O. and Martin, M. A. (1996). Domains of the human immunodeficiency virus type 1 matrix and gp41 cytoplasmic tail required for envelope incorporation into virions. *J. Virol.* **70**, 341-351.
- Freed, E. O. and Martin, M. A. (2001). HIVs and Their Replication. In "Fields Virology" (D. M. Knipe and P. M. Howley, Eds.), pp. 1971-2042. Lippincott Williams and Wilkins, Philadelphia.
- Fritz, C. C., Zapp, M. L., and Green, M. R. (1995). A human nucleoporin-like protein that specifically interacts with HIV rev. *Nature* **376**, 530-533.
- Fujiwara, T. and Mizuuchi, K. (1988). Retroviral DNA integration: structure of an integration intermediate. *Cell* **54**, 497-504.
- Furuta, R. A., Wild, C. T., Weng, Y., and Weiss, C. D. (1998). Capture of an early fusion-active conformation of HIV-1 gp41. *Nat. Struct. Biol.* **5**, 276-279.
- Gabuzda, D. and Wang, J. (2000). Chemokine receptors and mechanisms of cell death in HIV neuropathogenesis. *J. Neurovirol.* **6 Suppl 1:S24-32**, S24-32.
- Gabuzda, D. H., Lawrence, K., Langhoff, E., Terwilliger, E., Dorfman, T., Haseltine, W. A., and Sodroski, J. (1992). Role of vif in replication of human immunodeficiency virus type 1 in CD4+ T lymphocytes. *J. Virol.* **66**, 6489-6495.
- Gaines, H., von Sydow, M., Sonnerborg, A., Albert, J., Czajkowski, J., Pehrson, P. O., Chiodi, F., Moberg, L., Fenyo, E. M., Asjo, B., and . (1987). Antibody response in primary human immunodeficiency virus infection. *Lancet* **1**, 1249-1253.
- Gaines, H., von Sydow, M. A., von Stedingk, L. V., Biberfeld, G., Bottiger, B., Hansson, L. O., Lundbergh, P., Sonnerborg, A. B., Wasserman, J., and Strannegaard, O. O. (1990). Immunological changes in primary HIV-1 infection. *AIDS* **4**, 995-999.
- Gallay, P., Hope, T., Chin, D., and Trono, D. (1997). HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway. *Proc Natl Acad Sci USA* **94**, 9825-9830.
- Gallay, P., Swingler, S., Song, J. P., Bushman, F., and Trono, D. (1995). HIV nuclear import is governed by the phosphotyrosine- mediated binding of matrix to the core domain of integrase. *Cell* **83**, 569-576.
- Gallo, R. C., Salahuddin, S. Z., Popovic, M., Shearer, G. M., Kaplan, M., Haynes, B. F., Palker, T. J., Redfield, R., Oleske, J., Safai, B., and et al (1984). Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science.* **224**, 500-503.
- Gamble, T. R., Vajdos, F. F., Yoo, S. H., Worthylake, D. K., Houseweart, M., Sundquist, W. I., and Hill, C. P. (1996). Crystal structure of human cyclophilin A bound to the amino- terminal domain of HIV-1 capsid. *Cell* **87**, 1285-1294.
- Gamble, T. R., Yoo, S. H., Vajdos, F. F., vonSchwedler, U. K., Worthylake, D. K., Wang, H., McCutcheon, J. P., Sundquist, W. I., and Hill, C. P. (1997). Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science* **278**, 849-853.

- Gao, F., Bailes, E., Robertson, D. L., Chen, Y. L., Rodenburg, C. M., Michael, S. F., Cummins, L. B., Arthur, L. O., Peeters, M., Shaw, G. M., Sharp, P. M., and Hahn, B. H. (1999). Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* **397**, 436-441.
- Gao, F., Morrison, S. G., Robertson, D. L., Thornton, C. L., Craig, S., Karlsson, G., Sodroski, J., Morgado, M., Galvaocastro, B., Vonbriesen, H., Beddows, S., Weber, J., Sharp, P. M., Shaw, G. M., Hahn, B. H., Osmanov, S., Heyward, W. L., Esparza, J., Vandepierre, P., Karita, E., Sempala, S., Tugume, B., Biryahwaho, B., Wasi, C., Rubsamenwaigmann, H., Holmes, H., Newberry, A., Ranjbar, S., Tomlinson, P., Bradac, J., Mullins, J. I., Delwart, E. L., Cheingsongpopov, R., Kaleebu, P., Myers, G., Korber, B. T. M., Chipangwi, J., Taha, T., Desormeaux, J., Eiumtrakul, S., Natpratan, C., Khamboonruang, C., Miotti, P., Halsey, N. A., Vlahov, D., Nelson, K. E., Phair, J., Cao, Y., Moore, J. P., and et al. (1996a). Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. *J. Virol.* **70**, 1651-1667.
- Gao, F., Robertson, D. L., Carruthers, C. D., Li, Y. Y., Bailes, E., Kostrikis, L. G., Salminen, M. O., Bibollettruche, F., Peeters, M., Ho, D. D., Shaw, G. M., Sharp, P. M., and Hahn, B. H. (1998). An isolate of human immunodeficiency virus type 1 originally classified as subtype I represents a complex mosaic comprising three different group M subtypes (A, G, and I). *J. Virol.* **72**, 10234-10241.
- Gao, F., Robertson, D. L., Morrison, S. G., Hui, H. X., Craig, S., Decker, J., Fultz, P. N., Girard, M., Shaw, G. M., Hahn, B. H., and Sharp, P. M. (1996b). The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J. Virol.* **70**, 7013-7029.
- 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.
- Garcia, J. V. and Miller, A. D. (1991). Serine phosphorylation-independent down regulation of cell surface CD4 by nef. *Nature* **350**, 508-511.
- Gartner, S. (2000). HIV infection and dementia. *Science* **287**, 602-604.
- Gartner, S. and Liu, Y. (2002). Insights into the role of immune activation in HIV neuropathogenesis. *J. Neurovirol.* **8**, 69-75.
- Gartner, S., Markovits, P., Markovitz, D. M., Kaplan, H., Gallo, R. C., and Popovic, M. (1986). The role of mononuclear phagocytes in HTLV-III/LAV infection. *Science* **233**, 215-219.
- Gartner, S., McDonald, R. A., Hunter, E. A., Bouwman, F., Liu, Y., and Popovic, M. (1997). Gp120 sequence variation in brain and in T-lymphocyte human immunodeficiency virus type 1 primary isolates. *J. Hum. Virol.* **1**, 3-18.
- Gatanaga, H., Oka, S., Ida, S., Wakabayashi, T., Shioda, T., and Iwamoto, A. (1999). Active HIV-1 redistribution and replication in the brain with HIV encephalitis. *Arch Virol* **144**, 29-43.
- Geijtenbeek, T. B., Kwon, D. S., Torensma, R., van Vliet, S. J., van Duijnhoven, G. C., Middel, J., Cornelissen, I. L., Nottet, H. S., Kewalramani, V. N., Littman, D. R., Figdor, C. G., and van Kooyk, Y. (2000a). DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* **100**, 587-597.
- Geijtenbeek, T. B., Torensma, R., van Vliet, S. J., van Duijnhoven, G. C., Adema, G. J., van Kooyk, Y., and Figdor, C. G. (2000b). Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. *Cell* **100**, 575-585.
- Geraghty, R. J. and Panganiban, A. T. (1993). Human immunodeficiency virus type-1 vpu has a CD4-glycoprotein and an envelope glycoprotein-independent function. *J. Virol.* **67**, 4190-4194.

- Ghorpade, A., Nukuna, A., Che, M., Haggerty, S., Persidsky, Y., Carter, E., Carhart, L., Shafer, L., and Gendelman, H. E. (1998a). Human immunodeficiency virus neurotropism: an analysis of viral replication and cytopathicity for divergent strains in monocytes and microglia. *J Virol* **72**, 3340-3350.
- Ghorpade, A., Xia, M. Q., Hyman, B. T., Persidsky, Y., Nukuna, A., Bock, P., Che, M. H., Limoges, J., Gendelman, H. E., and Mackay, C. R. (1998b). Role of the beta-chemokine receptors CCR3 and CCR5 in human immunodeficiency virus type 1 infection of monocytes and microglia. *J Virol* **72**, 3351-3361.
- Gilboa, E., Mitra, S. W., Goff, S., and Baltimore, D. (1979). A detailed model of reverse transcription and tests of crucial aspects. *Cell* **18**, 93-100.
- Giulian, D., Vaca, K., and Noonan, C. A. (1990). Secretion of neurotoxins by mononuclear phagocytes infected with HIV-1. *Science* **250**, 1593-1596.
- Goff, S. P. (2001). *Retroviridae: The Retroviruses and Their Replication*. In "Fields Virology" (D. M. Knipe and P. M. Howley, Eds.), pp. 1871-1940. Lippincott Williams and Wilkins, Philadelphia.
- 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. *Journal of AIDS* **2**, 344-352.
- Gorry, P. R., Bristol, G., Zack, J. A., Ritola, K., Swannstrom, R., Birch, C. J., Bell, J. E., Bannert, N., Crawford, K., Wang, H., Schols, D., De Clercq, E., Kunstman, K., Wolinsky, S. M., and Gabuzda, D. (2001). Macrophage tropism of human immunodeficiency virus type 1 isolates from brain and lymphoid tissues predicts neurotropism independent of coreceptor specificity. *J. Virol.* **75**, 10073-10089.
- Gorry, P. R., Howard, J. L., Churchill, M. J., Anderson, J. L., Cunningham, A., Adrian, D., McPhee, D. A., and Purcell, D. F. (1999). Diminished production of human immunodeficiency virus type 1 in astrocytes results from inefficient translation of gag, env, and nef mRNAs despite efficient expression of Tat and Rev. *J. Virol.* **73**, 352-361.
- Gorry, P. R., Taylor, J., Holm, G. H., Mehle, A., Morgan, T., Cayabyab, M., Farzan, M., Wang, H., Bell, J. E., Kunstman, K., Moore, J. P., Wolinsky, S. M., and Gabuzda, D. (2002a). Increased CCR5 affinity and reduced CCR5/CD4 dependence of a neurovirulent primary human immunodeficiency virus type 1 isolate. *J. Virol.* **76**, 6277-6292.
- Gorry, P. R., Taylor, J., Holm, G. H., Mehle, A., Morgan, T., Cayabyab, M., Farzan, M., Wang, H., Bell, J. E., Kunstman, K., Moore, J. P., Wolinsky, S. M., and Gabuzda, D. (2002b). Increased CCR5 affinity and reduced CCR5/CD4 dependence of a neurovirulent primary human immunodeficiency virus type 1 isolate. *J. Virol.* **76**, 6277-6292.
- Gosling, J., Monteclaro, F. S., Atchison, R. E., Arai, H., Tsou, C. L., Goldsmith, M. A., and Charo, I. F. (1997). Molecular uncoupling of C-C chemokine receptor 5-induced chemotaxis and signal transduction from HIV-1 coreceptor activity. *Proc Natl Acad Sci USA* **94**, 5061-5066.
- Gottlinger, H., Sodroski, J., and Haseltine, W. (1989). Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type-1. *Proc. Natl. Acad. Sci. USA* **86**, 5781-5785.
- Goulder, P. J., Brander, C., Tang, Y., Tremblay, C., Colbert, R. A., Addo, M. M., Rosenberg, E. S., Nguyen, T., Allen, R., Trocha, A., Altfeld, M., He, S., Bunce, M., Funkhouser, R., Pelton, S. I., Burchett, S. K., McIntosh, K., Korber, B. T., and Walker, B. D. (2001). Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature* **412**, 334-338.

- Grattinger, M., Hohenberg, H., Thomas, D., Wilk, T., Muller, B., and Krausslich, H. G. (1999). In vitro assembly properties of wild-type and cyclophilin-binding defective human immunodeficiency virus capsid proteins in the presence and absence of cyclophilin A. *Virology* **257**, 247-260.
- Gray, F., Lescs, M. C., Keohane, C., Paraire, F., Marc, B., Durigon, M., and Gherardi, R. (1992). Early brain changes in HIV infection: neuropathological study of 11 HIV seropositive, non-AIDS cases. *J. Neuropathol. Exp. Neurol.* **51**, 177-185.
- Gray, F., Scaravilli, F., Everall, I., Chretien, F., An, S., Boche, D., Adle-Biassette, H., Wingertsmann, L., Durigon, M., Hurtrel, B., Chiodi, F., Bell, J., and Lantos, P. (1996). Neuropathology of early HIV-1 infection. *Brain Pathol.* **6**, 1-15.
- Greenberg, M. E., Bronson, S., Lock, M., Neumann, M., Pavlakis, G. N., and Skowronski, J. (1997). Co-localization of HIV-1 Nef with the AP-2 adaptor protein complex correlates with Nef-induced CD4 down-regulation. *EMBO J* **16**, 6964-6976.
- Gregory, R. J., Cheng, S. H., Rich, D. P., Marshall, J., Paul, S., Hehir, K., Ostedgaard, L., Klinger, K. W., Welsh, M. J., and Smith, A. E. (1990). Expression and characterization of the cystic fibrosis transmembrane conductance regulator. *Nature* **347**, 382-386.
- Grez, M., Dietrich, U., Balfe, P., Vonbriesen, H., Maniar, J. K., Mahambre, G., Delwart, E. L., Mullins, J. I., and Rubsamenwaigmann, H. (1994). 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**, 2161-2168.
- Groenink, M., Fouchier, R. A. M., Broersen, S., Baker, C. H., Koot, M., van't Wout, A. B., Huisman, H. G., Miedema, F., Tersmette, M., and Schuitemaker, H. (1993). Relation of phenotype evolution of HIV-1 to envelope V2 configuration. *Science* **260**, 1513-1516.
- Gunthard, H. F., Frost, S. D. W., LeighBrown, A. J., Ignacio, C. C., Kee, K., Perelson, A. S., Spina, C. A., Havlir, D. V., Hezareh, M., Looney, D. J., Richman, D. D., and Wong, J. K. (1999). Evolution of envelope sequences of human immunodeficiency virus type 1 in cellular reservoirs in the setting of potent antiviral therapy. *J. Virol* **73**, 9404-9412.
- Gunther, S., Sommer, G., Plikat, U., Iwanska, A., Wainhobson, S., Will, H., and Meyerhans, A. (1997). Naturally occurring hepatitis B virus genomes bearing the hallmarks of retroviral G->A hypermutation. *Virology* **235**, 104-108.
- Guo, J. H., Henderson, L. E., Bess, J., Kane, B., and Levin, J. G. (1997). Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA. *J. Virol* **71**, 5178-5188.
- Gyorkey, F., Melnick, J. L., and Gyorkey, P. (1987). Human immunodeficiency virus in brain biopsies of patients with AIDS and progressive encephalopathy. *Journal of Infectious Diseases* **155**, 870-876.
- Haggerty, S. and Stevenson, M. (1991). Predominance of distinct viral genotypes in brain and lymph node compartments of HIV-infected individuals. *Viral. Immunol.* **4**, 123-131.
- Hahn, B. H., Shaw, G. M., Arya, S. K., Popovic, M., Gallo, R. C., and Wong Staal, F. (1984). Molecular cloning and characterization of the HTLV-III virus associated with AIDS. *Nature.* **312**, 166-169.
- Hahn, B. H., Shaw, G. M., De Cock, K. M., and Sharp, P. M. (2000). AIDS as a zoonosis: scientific and public health implications. *Science* **287**, 607-614.

- Hahn, B. H., Shaw, G. M., Taylor, M. E., Redfield, R. R., Markham, P. D., Salahuddin, S. Z., Wong Staal, F., Gallo, R. C., Parks, E. S., and Parks, W. P. (1986). Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. *Science* **232**, 1548-1553.
- Hallenberger, S., Bosch, V., Angliker, H., Shaw, E., Klenk, H. D., and Garten, W. (1992). Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. *Nature* **360**, 358-361.
- Hammerschmid, M., Palmeri, D., Ruhl, M., Jaksche, H., Weichselbraun, I., Bohnlein, E., Malim, M. H., and Hauber, J. (1994). Scanning mutagenesis of the arginine-rich region of the human immunodeficiency virus type 1 rev trans activator. *J.Virol.* **68**, 7329-7335.
- Harouse, J. M., Bhat, S., Spitalnik, S. L., Laughlin, M., Stefano, K., Silberberg, D. H., and Gonzalez Scarano, F. (1991). Inhibition of entry of HIV-1 in neural cell lines by antibodies against galactosyl ceramide. *Science* **253**, 320-323.
- Harouse, J. M., Collman, R. G., and Gonzalezscarano, F. (1995). Human immunodeficiency virus type 1 infection of SK-n-MC cells: domains of gp120 involved in entry into a CD4- negative, galactosyl ceramide 3' sulfo-galactosyl ceramide- positive cell line. *J.Virol.* **69**, 7383-7390.
- Harris, C., Small, C. B., Klein, R. S., Friedland, G. H., Moll, B., Emeson, E. E., Spigland, I., and Steigbigel, N. H. (1983). Immunodeficiency in female sexual partners of men with the acquired immunodeficiency syndrome. *N.Engl.J.Med.* **308**, 1181-1184.
- Harris, R. S., Bishop, K. N., Sheehy, A. M., Craig, H. M., Petersen-Mahrt, S. K., Watt, I. N., Neuberger, M. S., and Malim, M. H. (2003). DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**, 803-809.
- Hart, A. R. and Cloyd, M. W. (1990). Interference patterns of human immunodeficiency viruses HIV-1 and HIV-2. *Virology* **177**, 1-10.
- He, J., Chen, Y., Farzan, M., Choe, H., Ohagen, A., Gartner, S., Busciglio, J., Yang, X., Hofmann, W., Newman, W., Mackay, C. R., Sodroski, J., and Gabuzda, D. (1997). CCR3 and CCR5 are co-receptors for HIV-1 infection of microglia. *Nature*. **385**, 645-649.
- Healey, D., Dianda, L., Moore, J. P., McDougal, J. S., Moore, M. J., Estess, P., Buck, D., Kwong, P. D., Beverley, P. C., and Sattentau, Q. J. (1990). Novel anti-CD4 monoclonal antibodies separate human immunodeficiency virus infection and fusion of CD4+ cells from virus binding. *J.Exp.Med.* **172**, 1233-1242.
- Heinzinger, N. K., Bukrinsky, M. I., Haggerty, S. A., Ragland, A. M., Kewalramani, V., Lee, M. A., Gendelman, H. E., Ratner, L., Stevenson, M., and Emerman, M. (1994). The vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. *Proc.Natl.Acad.Sci.USA* **91**, 7311-7315.
- Hellerstein, M. (2002). HIV tropism and CD4+ T-cell depletion. *Nat.Med.* **8**, 537-538.
- Henriksson, P. and Bosch, V. (1998). Inhibition of cellular glycoprotein incorporation into human immunodeficiency virus-like particles by coexpression of additional cellular interaction partner. *Virology* **251**, 16-21.
- Hesseltger, J. and Horuk, R. (1999). Chemokine and chemokine receptor expression in the central nervous system. *J.Neurovirol.* **5**, 13-26.
- Hibbitts, S., Reeves, J. D., Simmons, G., Gray, P. W., Epstein, L. G., Schols, D., De Clercq, E., Wells, T. N., Proudfoot, A. E., and Clapham, P. R. (1999). Coreceptor ligand inhibition of fetal brain cell infection by HIV type 1. *AIDS Res.Hum.Retroviruses* **15**, 989-1000.

- Hildreth, J. E. and Orentas, R. J. (1989). Involvement of a leukocyte adhesion receptor (LFA-1) in HIV- induced syncytium formation. *Science* **244**, 1075-1078.
- Hill, C. M., Kwon, D., Jones, M., Davis, C. B., Marmon, S., Daugherty, B. L., DeMartino, J. A., Springer, M. S., Unutmaz, D., and Littman, D. R. (1998). The amino terminus of human CCR5 is required for its function as a receptor for diverse human and simian immunodeficiency virus envelope glycoproteins. *Virology* **248**, 357-371.
- Hillis, D. M. and Bull, J. J. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst.Biol.* **42**, 182-192.
- Hirsch, V. M., Olmsted, R. A., Murphey Corb, M., Purcell, R. H., and Johnson, P. R. (1989). An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature*. **339**, 389-392.
- Ho, D. D., Neumann, A. U., Perelson, A. S., Chen, W., Leonard, J. M., and Markowitz, M. (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **373**, 123-126.
- Ho, D. D., Sarngadharan, M. G., Resnick, L., Dimarzo-veronese, F., Rota, T. R., and Hirsch, M. S. (1985). Primary human T-lymphotropic virus type III infection. *Ann.Intern.Med.* **103**, 880-883.
- Hoffman, T. L., Stephens, E. B., Narayan, O., and Doms, R. W. (1998). HIV type I envelope determinants for use of the CCR2b, CCR3, STRL33, and APJ coreceptors. *Proc Natl Acad Sci USA* **95**, 11360-11365.
- Holmes, E. C., Zhang, L. Q., Simmonds, P., Rogers, A. S., and Brown, A. J. L. (1993). Molecular investigation of human immunodeficiency virus (HIV) infection in a patient of an HIV-infected surgeon. *Journal of Infectious Diseases* **167**, 1411-1414.
- Hu, W. S., Bowman, E. H., Delviks, K. A., and Pathak, V. K. (1997). Homologous recombination occurs in a distinct retroviral subpopulation and exhibits high negative interference. *J Virol* **71**, 6028-6036.
- Hu, W. S. and Temin, H. (1990a). Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc.Natl.Acad.Sci.USA* **87**, 1556-1560.
- Hu, W. S. and Temin, H. M. (1990b). Retroviral recombination and reverse transcription. *Science* **250**, 1227-1233.
- Hu, W. S. and Temin, H. M. (1992). Effect of gamma radiation on retroviral recombination. *Journal of Virology* **66**, 4457-4463.
- Huang, H. F., Chopra, R., Verdine, G. L., and Harrison, S. C. (1998). Structure of a covalently trapped catalytic complex of HIV-I reverse transcriptase: Implications for drug resistance. *Science* **282**, 1669-1675.
- Huang, M. J., Orenstein, J. M., Martin, M. A., and Freed, E. O. (1995). p6 (gag) is required for particle production from full- length human immunodeficiency virus type 1 molecular clones expressing protease. *J.Virol.* **69**, 6810-6818.
- Huang, Y., Khorchid, A., Wang, J., Parniak, M. A., Darlix, J. L., Wainberg, M. A., and Kleiman, L. (1997). Effect of mutations in the nucleocapsid protein (NCp7) upon Pr160 (gag-pol) and tRNA (Lys) incorporation into human immunodeficiency virus type 1. *J Virol* **71**, 4378-4384.
- Huet, T., Cheyrier, R., Meyerhans, A., Roelants, G., and Wain Hobson, S. (1990). Genetic organization of a chimpanzee lentivirus related to HIV-1. *Nature* **345**, 356-359.

- Hughes, E. S., Bell, J. E., and Simmonds, P. (1997a). Investigation of population diversity of human immunodeficiency virus type 1 in vivo by nucleotide sequencing and length polymorphism analysis of the V1/V2 hypervariable region of env. *J. Gen. Virol.* **78**:2871-2882, 2871-2882.
- Hughes, E. S., Bell, J. E., and Simmonds, P. (1997b). Investigation of the dynamics of the spread of human immunodeficiency virus to brain and other tissues by evolutionary analysis of sequences from the p17 (gag) and env genes. *J. Virol.* **71**, 1272-1280.
- Hughes, M. D., Johnson, V. A., Hirsch, M. S., Bremer, J. W., Elbeik, T., Erice, A., Kuritzkes, D. R., Scott, W. A., Spector, S. A., Basgoz, N., Fischl, M. A., and DAquila, R. T. (1997c). Monitoring plasma HIV-1 RNA levels in addition to CD4 (+) lymphocyte count improves assessment of antiretroviral therapeutic response. *Ann Intern Med* **126**, 929-938.
- Hung, C. S., Vanderheyden, N., and Ratner, L. (1999). Analysis of the critical domain in the V3 loop of human immunodeficiency virus type 1 gp120 involved in CCR5 utilization. *J. Virol.* **73**, 8216-8226.
- Hunter, E. (1997). Viral Entry and Receptors. In "Retroviruses" (J. M. Coffin, S. E. Hughes, and H. E. Varmus, Eds.), pp. 71-120. Cold Spring Harbour Laboratory Press, New York.
- Hwang, S. S., Boyle, T. J., Lyster, H. K., and Cullen, B. R. (1991). Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science* **253**, 71-74.
- Imlach, S., Leen, C., Bell, J. E., and Simmonds, P. (2003). Phenotypic analysis of peripheral blood gammadelta T lymphocytes and their targeting by human immunodeficiency virus type 1 in vivo. *Virology* **305**, 415-427.
- Imlach, S., McBreen, S., Shirafuji, T., Leen, C., Bell, J. E., and Simmonds, P. (2001). Activated peripheral CD8 lymphocytes express CD4 in vivo and are targets for infection by human immunodeficiency virus type 1. *J. Virol.* **75**, 11555-64.
- Itescu, S., Simonelli, P. F., Winchester, R. J., and Ginsberg, H. S. (1994). Human immunodeficiency virus type 1 strains in the lungs of infected individuals evolve independently from those in peripheral blood and are highly conserved in the c- terminal region of the envelope v3 loop. *Proc. Natl. Acad. Sci. USA* **91**, 11378-11382.
- Ives, K. J., Jacobsen, H., Galpin, S. A., Garaev, M. M., Dorrell, L., Mous, J., Bragman, K., and Weber, J. N. (1997). Emergence of resistant variants of HIV in vivo during monotherapy with the proteinase inhibitor saquinavir. *J. Antimicrob. Chemother.* **39**, 771-779.
- Jacobomolina, A., Ding, J. P., Nanni, R. G., Clark, A. D., Lu, X. D., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., Hizi, A., Hughes, S. H., and Arnold, E. (1993). Crystal structure of human immunodeficiency virus type-1 reverse transcriptase complexed with double-stranded DNA at 3.0 angstrom resolution shows bent DNA. *Proc. Natl. Acad. Sci. USA* **90**, 6320-6324.
- Jacobsen, H., Yasargil, K., Winslow, D. L., Craig, J. C., Krohn, A., Duncan, I. B., and Mous, J. (1995). Characterization of human immunodeficiency virus type 1 mutants with decreased sensitivity to proteinase inhibitor ro 31-8959. *Virology* **206**, 527-534.
- Jaffe, H. W., Bregman, D. J., and Selik, R. M. (1983). Acquired immune deficiency syndrome in the United States: the first 1,000 cases. *Journal of Infectious Diseases* **148**, 339-345.
- Jahn, R. and Sudhof, T. C. (1999). Membrane fusion and exocytosis. *Annu. Rev. Biochem.* **68**, 863-911.
- Jenkins, G. M., Worobey, M., Woelk, C. H., and Holmes, E. C. (2001). Evidence for the non-quasispecies evolution of RNA viruses [corrected]. *Mol. Biol. Evol.* **18**, 987-994.

- Jetzt, A. E., Yu, H., Klarmann, G. J., Ron, Y., Preston, B. D., and Dougherty, J. P. (2000). High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J. Virol.* **74**, 1234-1240.
- Ji, J. P. and Loeb, L. A. (1992). Fidelity of HIV-1 reverse transcriptase copying RNA in vitro. *Biochemistry* **31**, 954-958.
- Ji, J. P. and Loeb, L. A. (1994). Fidelity of HIV-1 reverse transcriptase copying a hypervariable region of the HIV-1 env gene. *Virology* **199**, 323-330.
- Jiang, M., Mak, J., Ladha, A., Cohen, E., Klein, M., Rovinski, B., and Kleiman, L. (1993). Identification of tRNAs incorporated into wild-type and mutant human immunodeficiency virus type-1. *J. Virol.* **67**, 3246-3253.
- Jones, J. S., Allan, R. W., and Temin, H. M. (1994). One retroviral RNA is sufficient for synthesis of viral DNA. *J. Virol.* **68**, 207-216.
- Judo, M. S. B., Wedel, A. B., and Wilson, C. (1998). Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Research* **26**, 1819-1825.
- Jukes, Y. H. and Cantor, C. R. (1969). Evolution of protein molecules. In "Mammalian protein metabolism." (H. N. Munro, Ed.), pp. 21-132. Academic Press, New York.
- Julias, J. G., Kim, T., Arnold, G., and Pathak, V. K. (1997). The antiretrovirus drug 3'-azido-3'-deoxythymidine increases the retrovirus mutation rate. *J. Virol.* **71**, 4254-4263.
- Jung, A., Maier, R., Vartanian, J. P., Bocharov, G., Jung, V., Fischer, U., Meese, E., Wain-Hobson, S., and Meyerhans, A. (2002). Multiply infected spleen cells in HIV patients. *Nature* **418**, 144.
- Junghans, R. P., Boone, L. R., and Skalka, A. M. (1982). Retroviral DNA H structures: displacement-assimilation model of recombination. *Cell* **30**, 53-62.
- Kabat, D., Kozak, S. L., Wehrly, K., and Chesebro, B. (1994). Differences in CD4 dependence for infectivity of laboratory-adapted and primary patient isolates of human immunodeficiency virus type 1. *J. Virol.* **68**, 2570-2577.
- Kanki, P. J., Travers, K. U., Mboup, S., Hsieh, C. C., Marlink, R. G., Gueyendiaeye, A., Siby, T., Thior, I., Hernandezavila, M., Sankale, J. L., Ndoye, I., and Essex, M. E. (1994). Slower heterosexual spread of HIV-2 than HIV-1. *Lancet* **343**, 943-946.
- Kanmogne, G. D., Kennedy, R. C., and Grammas, P. (2002). HIV-1 gp120 proteins and gp160 peptides are toxic to brain endothelial cells and neurons: possible pathway for HIV entry into the brain and HIV-associated dementia. *J. Neuropathol. Exp. Neurol.* **61**, 992-1000.
- Kasper, P., Simmonds, P., Schneweis, K. E., Kaiser, R., Matz, B., Oldenburg, J., Brackmann, H. H., and Holmes, E. C. (1995). The genetic diversification of the HIV type 1 gag p17 gene in patients infected from a common source. *AIDS Res. Hum. Retroviruses* **11**, 1197-1201.
- Kato, K., Sato, H., and Takebe, Y. (1999). Role of naturally occurring basic amino acid substitutions in the human immunodeficiency virus type 1 subtype E envelope V3 loop on viral coreceptor usage and cell tropism. *J. Virol.* **73**, 5520-5526.
- Katz, R. A. and Skalka, A. M. (1990). Generation of diversity in retroviruses. *Annu. Rev. Genet.* **24**, 409-445.
- Kaul, M., Garden, G. A., and Lipton, S. A. (2001). Pathways to neuronal injury and apoptosis in HIV-associated dementia. *Nature* **410**, 988-994.

- Kaul, M. and Lipton, S. A. (1999). Chemokines and activated macrophages in HIV gp120-induced neuronal apoptosis. *Proc Nat Acad Sci Usa* **96**, 8212-8216.
- Kawai, S. and Hanafusa, H. (1972). Genetic recombination with avian tumor virus. *Virology* **49**, 37-44.
- Kawakami, T., Sherman, L., Dahlberg, J., Gazit, A., Yaniv, A., Tronick, S. R., and Aaronson, S. A. (1987). Nucleotide sequence analysis of equine infectious anemia virus proviral DNA. *Virology* **158**, 300-312.
- Kellam, P. and Larder, B. A. (1995). Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance. *J. Virol.* **69**, 669-674.
- Kiernan, R. E., Ono, A., Englund, G., and Freed, E. O. (1998). Role of matrix in an early postentry step in the human immunodeficiency virus type 1 life cycle. *J Virol* **72**, 4116-4126.
- Kilby, J. M., Hopkins, S., Venetta, T. M., DiMassimo, B., Cloud, G. A., Lee, J. Y., Alldredge, L., Hunter, E., Lambert, D., Bolognesi, D., Mathews, T., Johnson, M. R., Nowak, M. A., Shaw, G. M., and Saag, M. S. (1998). Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. *Nature Med* **4**, 1302-1307.
- Kim, J. K., Palaniappan, C., Wu, W. M., Fay, P. J., and Bambara, R. A. (1997). Evidence for a unique mechanism of strand transfer from the transactivation response region of HIV-1. *J Biol Chem* **272**, 16769-16777.
- Kim, S., Ikeuchi, K., Groopman, J., and Baltimore, D. (1990). Factors affecting cellular tropism of Human immunodeficiency virus. *J. Virol.* **64**, 5600-5604.
- Kim, S. Y., Byrn, R., Groopman, J., and Baltimore, D. (1989). Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. *J. Virol.* **63**, 3708-3713.
- Kim, T., Mudry, R. A., Rexrode, C. A., and Pathak, V. K. (1996). Retroviral mutation rates and A-to-G hypermutations during different stages of retroviral replication. *J. Virol.* **70**, 7594-7602.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111-120.
- King, W., Patel, M. D., Lobel, L. I., Goff, S. P., and Nguyen-Huu, M. C. (1985). Insertion mutagenesis of embryonal carcinoma cells by retroviruses. *Science* **228**, 554-558.
- Kinter, A., Arthos, J., Cicala, C., and Fauci, A. S. (2000). Chemokines, cytokines and HIV: a complex network of interactions that influence HIV pathogenesis. *Immunol. Rev.* **177**, 88-98.
- Kitamura, Y., Lee, Y. M., and Coffin, J. M. (1992). Nonrandom integration of retroviral DNA in vitro: effect of CpG methylation. *Proc. Natl. Acad. Sci. U.S.A* **89**, 5532-5536.
- Klarmann, G. J., Schaubert, C. A., and Preston, B. D. (1993). Template-directed pausing of DNA synthesis by HIV-1 reverse transcriptase during polymerization of HIV-1 sequences invitro. *Journal of Biological Chemistry* **268**, 9793-9802.
- Klasens, B. I., Huthoff, H. T., Das, A. T., Jeeninga, R. E., and Berkhout, B. (1999). The effect of template RNA structure on elongation by HIV-1 reverse transcriptase. *Biochim. Biophys. Acta* **1444**, 355-370.
- Klatzmann, D., Barre-Sinoussi, F., and Nugeyre, M. T. (1984a). Selective tropism of lymphadenopathy-associated virus (LAV) for helper induced lymphocytes. *Science* **225**, 59-63.

- Klatzmann, D., Champagne, E., Chamaret, S., Gruet, J., Guetard, D., Hercend, T., Gluckman, J. C., and Montagnier, L. (1984b). T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature*. **312**, 767-768.
- Koenig, S., Gendelman, H. E., Orenstein, J. M., Dal Canto, M. C., Pezeshkpoor, G. H., Yungbluth, M., Janotta, F., Aksamit, A., Martin, M. A., and Fauci, A. S. (1986). Detection of AIDS virus in macrophages in brain tissue from AIDS patients with encephalopathy. *Science*. **233**, 1089-1093.
- Koito, A., Harrowe, G., Levy, J. A., and Chengmayer, C. (1994). Functional role of the V1/V2 region of human immunodeficiency virus type 1 envelope glycoprotein gp120 in infection of primary macrophages and soluble CD4 neutralization. *J.Virol.* **68**, 2253-2259.
- Koito, A., Stamatatos, L., and Cheng-Mayer, C. (1995). Small amino acid sequence changes within the V2 domain can affect the function of a T cell line tropic human immunodeficiency virus type 1 envelope gp120. *Virology*. **206**, 878-884.
- Kolchinsky, P., Kiprilov, E., Bartley, P., Rubinstein, R., and Sodroski, J. (2001a). Loss of a single N-linked glycan allows CD4-independent human immunodeficiency virus type 1 infection by altering the position of the gp120 V1/V2 variable loops. *J.Virol.* **75**, 3435-3443.
- Kolchinsky, P., Kiprilov, E., and Sodroski, J. (2001b). Increased neutralization sensitivity of CD4-independent human immunodeficiency virus variants. *J.Virol.* **75**, 2041-2050.
- Kolchinsky, P., Mirzabekov, T., Farzan, M., Kiprilov, E., Cayabyab, M., Mooney, L. J., Choe, H., and Sodroski, J. (1999). Adaptation of a CCR5-using, primary human immunodeficiency virus type 1 isolate for CD4-independent replication. *J.Virol.* **73**, 8120-8126.
- Kondo, E. and Gottlinger, H. G. (1996). A conserved LXXLF sequence is the major determinant in p6 (gag) required for the incorporation of human immunodeficiency virus type 1 Vpr. *J.Virol.* **70**, 159-164.
- Koot, M., Keet, I. P. M., Vos, A. H. V., Degoede, R. E. Y., Roos, M. T. L., Coutinho, R. A., Miedema, F., Schellekens, P. T. A., and Tersmette, M. (1993). Prognostic value of HIV-1 syncytium-inducing phenotype for rate of CD4+ cell depletion and progression to AIDS. *Ann.Intern.Med.* **118**, 681-688.
- Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B. H., Wolinsky, S., and Bhattacharya, T. (2000). Timing the ancestor of the HIV-1 pandemic strains. *Science* **288**, 1789-1796.
- Korber, B. T. M., Kunstman, K. J., Patterson, B. K., Furtado, M., Mcevilly, M. M., Levy, R., and Wolinsky, S. M. (1994a). 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.
- Korber, B. T. M., Macinnes, K., Smith, R. F., and Myers, G. (1994b). Mutational trends in v3 loop protein sequences observed in different genetic lineages of human immunodeficiency virus type 1. *J.Virol.* **68**, 6730-6744.
- Kostrikis, L. G., Bagdades, E., Cao, Y. Z., Zhang, L. Q., Dimitriou, D., and Ho, D. D. (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.
- Koup, R. A., Safrit, J. T., Cao, Y. Z., Andrews, C. A., Mcleod, G., Borkowsky, W., Farthing, C., and Ho, D. D. (1994). Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *J.Virol.* **68**, 4650-4655.

- Kozak, S. L., Platt, E. J., Madani, N., Ferro, F. E., Peden, K., and Kabat, D. (1997). CD4, CXCR-4, and CCR-5 dependencies for infections by primary and laboratory-adapted isolates of human immunodeficiency virus type 1. *J. Virol.* **71**, 873-882.
- Kumar, S., Tamura, k., and Nei, M. (1993). "MEGA: Molecular Evolutionary Genetics Analysis, version 1.0." The Pennsylvania State University., University Park.
- Kwok, S., Kellogg, D. E., Mckinney, N., Spasic, D., Goda, L., Levenson, C., and Sninsky, J. J. (1990). Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Research* **18**, 999-1005.
- Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648-659.
- Labranche, C. C., Hoffman, T. L., Romano, J., Haggarty, B. S., Edwards, T. G., Matthews, T. J., Doms, R. W., and Hoxie, J. A. (1999). Determinants of CD4 independence for a human immunodeficiency virus type 1 variant map outside regions required for coreceptor specificity. *J Virol* **73**, 10310-10319.
- Lackritz, E. M., Satten, G. A., Aberlegrasse, J., Dodd, R. Y., Raimondi, V. P., Janssen, R. S., Lewis, W. F., Notari, E. P., and Petersen, L. R. (1995). Estimated risk of transmission of the human immunodeficiency virus by screened blood in the United States. *N.Engl.J.Med* **333**, 1721-1725.
- Lapham, C. K., Zaitseva, M. B., Lee, S., Romanstseva, T., and Golding, H. (1999). Fusion of monocytes and macrophages with HIV-1 correlates with biochemical properties of CXCR4 and CCR5. *Nature Med* **5**, 303-308.
- Larder, B. A. and Kemp, S. D. (1989). Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science.* **246**, 1155-1158.
- Larkin, M., Childs, R. A., Matthews, T. J., Thiel, S., Mizuochi, T., Lawson, A. M., Savill, J. S., Haslett, C., Diaz, R., and Feizi, T. (1989). Oligosaccharide-mediated interactions of the envelope glycoprotein gp120 of HIV-1 that are independent of CD4 recognition. *AIDS* **3**, 793-798.
- Lathey, J. L., Pratt, R. D., and Spector, S. A. (1997). Appearance of autologous neutralizing antibody correlates with reduction in virus load and phenotype switch during primary infection with human immunodeficiency virus type 1. *J Infect Dis* **175**, 231-232.
- Laughrea, M., Jette, L., Mak, J., Kleiman, L., Liang, C., and Wainberg, M. A. (1997). Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization. *J Virol* **71**, 3397-3406.
- Lecossier, D., Bouchonnet, F., Clavel, F., and Hance, A. J. (2003). Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science* **300**, 1112.
- Lee, B., Sharron, M., Blanpain, C., Doranz, B. J., Vakili, J., Setoh, P., Berg, E., Liu, G., Guy, H. R., Durell, S. R., Parmentier, M., Chang, C. N., Price, K., Tsang, M., and Doms, R. W. (1999a). Epitope mapping of CCR5 reveals multiple conformational states and distinct but overlapping structures involved in chemokine and coreceptor function. *J Biol Chem* **274**, 9617-9626.
- Lee, B., Sharron, M., Montaner, L. J., Weissman, D., and Doms, R. W. (1999b). Quantification of CD4, CCR5, and CXCR4 levels on lymphocyte subsets, dendritic cells, and differentially conditioned monocyte-derived macrophages. *Proc.Natl.Acad.Sci.U.S.A* **96**, 5215-5220.

- Lee, S., Tiffany, H. L., King, L., Murphy, P. M., Golding, H., and Zaitseva, M. B. (2000). CCR8 on human thymocytes functions as a human immunodeficiency virus type 1 coreceptor. *J. Virol.* **74**, 6946-6952.
- Lenburg, M. E. and Landau, N. R. (1993). Vpu-induced degradation of CD4 - requirement for specific amino acid residues in the cytoplasmic domain of CD4. *J. Virol.* **67**, 7238-7245.
- Leonard, C. K., Spellman, M. W., Riddle, L., Harris, R. J., Thomas, J. N., and Gregory, T. J. (1990). Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *Journal of Biological Chemistry* **265**, 10373-10382.
- Levy, J. A., Hoffman, A. D., Kramer, S. M., Landis, J. A., and Shimabukuro, J. M. (1984). Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science* **225**, 840-842.
- Lewin, S. R., Sonza, S., Irving, L. B., McDonald, C. F., Mills, J., and Crowe, S. M. (1996). Surface CD4 is critical to in vitro HIV infection of human alveolar macrophages. *AIDS Res. Hum. Retroviruses* **12**, 877-883.
- Lewis, L. L., Venzon, D., Church, J., Farley, M., Wheeler, S., Keller, A., Rubin, M., Yuen, G., Mueller, B., Sloas, M., Wood, L., Balis, F., Shearer, G. M., Brouwers, P., Goldsmith, J., Pizzo, P. A., Anderson, B., Hirschfeld, S., Sei, S., Zeichner, S., Roilides, E., Clerici, M., Wells, M., Stocker, V., Higham, C., Townley, E., Walsek, C., Zwierski, S., Heilman, N., SelkinGutman, G., Wolters, P., Moss, H., and Jarosinski, P. (1996). Lamivudine in children with human immunodeficiency virus infection: A phase I/II study. *Journal of Infectious Diseases* **174**, 16-25.
- Li, C. J., Friedman, D. J., Wang, C. L., Metelev, V., and Pardee, A. B. (1995). Induction of apoptosis in uninfected lymphocytes by HIV-1 tat protein. *Science* **268**, 429-431.
- Li, W. H., Tanimura, M., and Sharp, P. M. (1988). Rates and dates of divergence between AIDS virus nucleotide sequences. *Mol. Biol. Evol.* **5**, 313-330.
- Li, X. G., Quan, Y. D., Arts, E. J., Li, Z., Preston, B. D., Derocquigny, H., Roques, B. P., Darlix, J. L., Kleiman, L., Parniak, M. A., and Wainberg, M. A. (1996). Human immunodeficiency virus type 1 nucleocapsid protein (NCp7) directs specific initiation of minus-strand DNA synthesis primed by human tRNA (3) (Lys) in vitro: Studies of viral RNA molecules mutated in regions that flank the primer binding site. *Journal of Virology* **70**, 4996-5004.
- Li, Y. X., 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 uncultured human brain tissue: identification of replication-competent and -defective viral genomes. *J. Virol.* **65**, 3973-3985.
- Libby, R. T. and Gallant, J. A. (1991). The role of RNA polymerase in transcriptional fidelity. *Mol. Microbiol.* **5**, 999-1004.
- Liitsola, K., Ristola, M., Holmstrom, P., Salminen, M., Brummer-Korvenkontio, H., Simola, S., Suni, J., and Leinikki, P. (2000). An outbreak of the circulating recombinant form AECM240 HIV-1 in the Finnish injection drug user population. *AIDS* **14**, 2613-2615.
- Liitsola, K., Tashkinova, I., Laukkanen, T., Korovina, G., Smolskaja, T., Momot, O., Mashkilleysen, N., Chaplinskias, S., Brummerkorvenkontio, H., Vanhatalo, J., Leinikki, P., and Salminen, M. O. (1998). HIV-1 genetic subtype A/B recombinant strain causing an explosive epidemic in injecting drug users in Kaliningrad. *AIDS* **12**, 1907-1919.
- Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715.

- Lipton, S. A. (1993). Human immunodeficiency virus-infected macrophages, gp120, and N-methyl-D-aspartate receptor-mediated neurotoxicity. *Ann.Neurol.* **33**, 227-228.
- Lipton, S. A. and Gendelman, H. E. (1995). Dementia associated with the acquired immunodeficiency syndrome. *N.Engl.J.Med.* **332**, 934-940.
- Little, S. J., Riggs, N. L., Chowers, M. Y., Fitch, N. J. S., Richman, D. D., Spina, C. A., and Guatelli, J. C. (1994). Cell surface CD4 downregulation and resistance to superinfection induced by a defective provirus of HIV-1. *Virology* **205**, 578-582.
- Liu, S. L., Mittler, J. E., Nickle, D. C., Mulvania, T. M., Shriner, D., Rodrigo, A. G., Kosloff, B., He, X., Corey, L., and Mullins, J. I. (2002). Selection for human immunodeficiency virus type 1 recombinants in a patient with rapid progression to AIDS. *J Virol* **76**, 10674-10684.
- Liu, S. L., Rodrigo, A. G., Shankarappa, R., and Learn, G. H. (1996). HIV quasispecies and resampling. *Science* **273**, 415-416.
- Liu, S. L., Schacker, T., Musey, L., Shriner, D., McElrath, M. J., Corey, L., and Mullins, J. I. (1997). Divergent patterns of progression to AIDS after infection from the same source: Human immunodeficiency virus type 1 evolution and antiviral responses. *J Virol* **71**, 4284-4295.
- Liu, Y., Tang, X. P., McArthur, J. C., Scott, J., and Gartner, S. (2000). Analysis of human immunodeficiency virus type 1 gp160 sequences from a patient with HIV dementia: evidence for monocyte trafficking into brain. *J.Neurovirol.* **6 Suppl 1:S70-81**, S70-81.
- Louwagie, J., Janssens, W., Mascola, J., Heyndrickx, L., Hegerich, P., Vandergroen, G., McCutchan, F. E., and Burke, D. S. (1995). Genetic diversity of the envelope glycoprotein from human immunodeficiency virus type 1 isolates of african origin. *J.Virol.* **69** , 263-271.
- Louwagie, J., Mccutchan, F., Mascola, J., Eddy, G., Fransen, K., Peeters, M., Vandergroen, G., and Burke, D. (1993). Genetic subtypes of HIV-1. *AIDS Res.Hum.Retroviruses* **9**, S147-S150.
- Loveday, C., Kaye, S., Tenantflowers, M., Semple, M., Ayliffe, U., Weller, I. V. D., and Tedder, R. S. (1995). HIV-1 RNA serum-load and resistant viral genotypes during early zidovudine therapy. *Lancet* **345**, 820-824.
- Lu, Z. H., Berson, J. F., Chen, Y. H., Turner, J. D., Zhang, T. Y., Sharron, M., Jenks, M. H., Wang, Z. X., Kim, J., Rucker, J., Hoxie, J. A., Peiper, S. C., and Doms, R. W. (1997). Evolution of HIV-1 coreceptor usage through interactions with distinct CCR5 and CXCR4 domains. *Proc Natl Acad Sci USA* **94**, 6426-6431.
- Ludwig, E., CeccheriniSilberstein, F., vanEmpel, J., Erfle, V., Neumann, M., and Brackwerner, R. (1999). Diminished Rev-mediated stimulation of human immunodeficiency virus type 1 protein synthesis is a hallmark of human astrocytes. *J Virol* **73**, 8279-8289.
- Lukashov, V. V., Kuiken, C. L., and Goudsmit, J. (1995). Intrahost human immunodeficiency virus type 1 evolution is related to length of the immunocompetent period. *J.Virol.* **69**, 6911-6916.
- Luo, G. X. and Taylor, J. (1990). Template switching by reverse transcriptase during DNA synthesis. *J.Virol.* **64**, 4321-4328.
- Luster, A. D. (1998). Mechanisms of disease: Chemokines - Chemotactic cytokines that mediate inflammation. *N Engl J Med* **338**, 436-445.
- Ly, A. and Stamatatos, L. (2000). V2 loop glycosylation of the human immunodeficiency virus type 1 SF162 envelope facilitates interaction of this protein with CD4 and CCR5 receptors and protects the

- virus from neutralization by anti-V3 loop and anti-CD4 binding site antibodies. *J. Virol.* **74**, 6769-6776.
- Macatonia, S. E., Lau, R., Patterson, S., Pinching, A. J., and Knight, S. C. (1990). Dendritic cell infection, depletion and dysfunction in HIV-infected individuals. *Immunology* **71**, 38-45.
- Maddon, P. J., Dalgleish, A. G., McDougal, J. S., Clapham, P. R., Weiss, R. A., and Axel, R. (1986). The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell.* **47**, 333-348.
- Major, E. O., Rausch, D., Marra, C., and Clifford, D. (2000). HIV-associated dementia. *Science* **288** , 440-442.
- Maldarelli, F., Chen, M. Y., Willey, R. L., and Strebel, K. (1993). Human immunodeficiency virus type-1 vpu protein is an oligomeric type-i integral membrane protein. *J. Virol.* **67** , 5056-5061.
- Malim, M. H., Hauber, J., Le, S. Y., Maizel, J. V., and Cullen, B. R. (1989). The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* **338**, 254-257.
- Mangasarian, A. and Trono, D. (1997). The multifaceted role of HIV Nef. *Res Virology* **148**, 30-33.
- Mansky, L. M. (1996). The mutation rate of human immunodeficiency virus type 1 is influenced by the vpr gene. *Virology* **222**, 391-400.
- Mansky, L. M. (1998). Retrovirus mutation rates and their role in genetic variation. *J Gen Virol* **79**, 1337-1345.
- Mansky, L. M. and Bernard, L. C. (2000). 3'-Azido-3'-deoxythymidine (AZT) and AZT-resistant reverse transcriptase can increase the in vivo mutation rate of human immunodeficiency virus type 1. *J. Virol.* **74**, 9532-9539.
- Mansky, L. M., Pearl, D. K., and Gajary, L. C. (2002). Combination of drugs and drug-resistant reverse transcriptase results in a multiplicative increase of human immunodeficiency virus type 1 mutant frequencies. *J Virol* **76**, 9253-9259.
- Mansky, L. M., Preveral, S., Selig, L., Benarous, R., and Benichou, S. (2000). The interaction of vpr with uracil DNA glycosylase modulates the human immunodeficiency virus type 1 In vivo mutation rate. *J. Virol.* **74**, 7039-7047.
- Mansky, L. M. and Temin, H. M. (1995). Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J. Virol.* **69**, 5087-5094.
- Marlink, R., Kanki, P., Thior, I., Travers, K., Eisen, G., Siby, T., Traore, I., Hsieh, C. C., Dia, M. C., Gueye, E., Hellinger, J., Gueyendiaye, A., Sankale, J. L., Ndoeye, I., Mboup, S., and Essex, M. (1994). Reduced rate of disease development after HIV-2 infection as compared to HIV-1. *Science* **265**, 1587-1590.
- Marmor, M., Friedman-Kien, A. E., Laubenstein, L., Byrum, R. D., William, D. C., D'onofrio, S., and Dubin, N. (1982). Risk factors for Kaposi's sarcoma in homosexual men. *Lancet* **1**, 1083-1087.
- Marras, D., Bruggeman, L. A., Gao, F., Tanji, N., Mansukhani, M. M., Cara, A., Ross, M. D., Gusella, G. L., Benson, G., D'Agati, V. D., Hahn, B. H., Klotman, M. E., and Klotman, P. E. (2002). Replication and compartmentalization of HIV-1 in kidney epithelium of patients with HIV-associated nephropathy. *Nat. Med.* **8**, 522-526.

- Martell, M., Esteban, J. I., Quer, J., Genesca, J., Weiner, A., Esteban, R., Guardia, J., and Gomez, J. (1992). Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. *J. Virol.* **66**, 3225-3229.
- Martell, M., Esteban, J. I., Quer, J., Vargas, V., Esteban, R., Guardia, J., and Gomez, J. (1994). Dynamic behavior of hepatitis C virus quasispecies in patients undergoing orthotopic liver transplantation. *J. Virol.* **68**, 3425-3436.
- Martin, I., Ruyschaert, J. M., and Epand, R. M. (1999). Role of the N-terminal peptides of viral envelope proteins in membrane fusion. *Advan Drug Delivery Rev* **38**, 233-255.
- Martin, J., Labranche, C. C., and Gonzalez-Scarano, F. (2001). Differential CD4/CCR5 utilization, gp120 conformation, and neutralization sensitivity between envelopes from a microglia- adapted human immunodeficiency virus type 1 and its parental isolate. *J. Virol.* **75**, 3568-80.
- Martinez, M. A., Vartanian, J. P., and Wainhobson, 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.
- Maynard Smith, J. (1978). "The evolution of sex." Cambridge University Press, Cambridge.
- McArthur, J. C., Hoover, D. R., Bacellar, H., Miller, E. N., Cohen, B. A., Becker, J. T., Graham, N. M., McArthur, J. H., Selnes, O. A., Jacobson, L. P., and . (1993). Dementia in AIDS patients: incidence and risk factors. Multicenter AIDS Cohort Study. *Neurology* **43**, 2245-2252.
- McBreen, S., Imlach, S., Scott, G. R., Leen, C., Bell, J. E., and Simmonds, P. (2001). Preferential infection of the CD45RA+ (naive) subset of CD8+ lymphocytes by human immunodeficiency virus type 1 in vivo. *J. Virol.* **75**, 4091-4102.
- McClure, P., Curran, R., Boneham, S., and Ball, J. K. (2000). A polymerase chain reaction method for the amplification of full-length envelope genes of HIV-1 from DNA samples containing single molecules of HIV-1 provirus. *J. Virol. Methods* **88**, 73-80.
- McCune, J. M. (2001). The dynamics of CD4+ T-cell depletion in HIV disease. *Nature* **410**, 974-979.
- McCutchan, F. E., Carr, J. K., Bajani, M., Sandersbuell, E., Harry, T. O., Stoeckli, T. C., Robbins, K. E., Gashau, W., Nasidi, A., Janssens, W., and Kalish, M. L. (1999). Subtype G and multiple forms of A/G intersubtype recombinant human immunodeficiency virus type 1 in Nigeria. *Virology* **254**, 226-234.
- McDougal, J. S., Kennedy, M. S., Sligh, J. M., Cort, S., Mawle, A. C., and Nicholson, J. K. A. (1986). Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. *Science* **231**, 382-385.
- McKnight, A., Dittmar, M. T., MonizPeriera, J., Ariyoshi, K., Reeves, J. D., Hibbitts, S., Whitby, D., Aarons, E., Proudfoot, A. E. I., Whittle, H., and Clapham, P. R. (1998). A broad range of chemokine receptors are used by primary isolates of human immunodeficiency virus type 2 as coreceptors with CD4. *J. Virol* **72**, 4065-4071.
- McNearney, T., Hornickova, Z., Markham, R., Birdwell, A., Arens, M., Saah, A., and Ratner, L. (1992). Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10247-10251.
- Mellado, M., RodriguezFrade, J. M., VilaCoro, A. J., deAna, A. M., and Martinez, C. (1999). Chemokine control of HIV-1 infection. *Nature* **400**, 723-724.

- Meyer, P. R., Matsuura, S. E., So, R. G., and Scott, W. A. (1998). Unblocking of chain-terminated primer by HIV-1 reverse transcriptase through a nucleotide-dependent mechanism. *Proc Natl Acad Sci USA* **95**, 13471-13476.
- Meyerhans, A., Cheynier, R., Albert, J., Seth, M., Kwok, S., Sninsky, J., Morfeldt Manson, L., Asjo, B., and Wain Hobson, S. (1989). Temporal fluctuations in HIV quasiespecies in vivo are not reflected by sequential HIV isolations. *Cell* **58**, 901-910.
- Meyerhans, A., Vartanian, J. P., and Wain Hobson, S. (1990). DNA recombination during PCR. *Nucleic Acids Research* **18**, 1687-1691.
- Michael, N. L. and Moore, J. P. (1999). HIV-1 entry inhibitors: Evading the issue. *Nature Med* **5**, 740-742.
- Mikkelsen, J. G. and Pedersen, F. S. (2000). Genetic reassortment and patch repair by recombination in retroviruses. *J. Biomed. Sci.* **7**, 77-99.
- Miller, M. D., Wang, B. B., and Bushman, F. D. (1995). Human immunodeficiency virus type 1 preintegration complexes containing discontinuous plus strands are competent to integrate in vitro. *J. Virol.* **69**, 3938-3944.
- Mizuno, A., Ido, E., Goto, T., Kuwata, T., Nakai, M., and Hayami, M. (1996). Mutational analysis of two zinc finger motifs in HIV type 1 nucleocapsid proteins: Effects on proteolytic processing of Gag precursors and particle formation. *AIDS Res. Hum. Retroviruses* **12**, 793-800.
- Moir, S., Perreault, J., and Poulin, L. (1996). Postbinding events mediated by human immunodeficiency virus type 1 are sensitive to modifications in the D4-transmembrane linker region of CD4. *J. Virol.* **70**, 8019-8028.
- Mondor, I., Ugolini, S., and Sattentau, Q. J. (1998). Human immunodeficiency virus type 1 attachment to HeLa CD4 cells is CD4 independent and gp120 dependent and requires cell surface heparans. *J. Virol* **72**, 3623-3634.
- Montavon, C., Bibollet-Ruche, F., Robertson, D., Koumare, B., Mulanga, C., Esu-Williams, E., Toure, C., Mboup, S., Soman, E., Delaporte, E., and Peeters, M. (1999). The identification of a complex A/G/I/J recombinant HIV type 1 virus in various West African countries. *AIDS Res. Hum. Retroviruses* **15**, 1707-1712.
- Montavon, C., Vergne, L., Bourgeois, A., Mpoudi-Ngole, E., Malonga-Mouellet, G., Butel, C., Toure-kane, C., Delaporte, E., and Peeters, M. (2002). Identification of a new circulating recombinant form of HIV type 1, CRF11-cpx, involving subtypes A, G, J, and CRF01-AE, in Central Africa. *AIDS Res. Hum. Retroviruses* **18**, 231-236.
- Montefiori, D. C., Pantaleo, G., Fink, L. M., Zhou, J. T., Zhou, J. Y., Bilska, M., Miralles, G. D., and Fauci, A. S. (1996). Neutralizing and infection-enhancing antibody responses to human immunodeficiency virus type 1 in long-term nonprogressors. *Journal of Infectious Diseases* **173**, 60-67.
- Moore, J. P., Sattentau, Q. J., Wyatt, R., and Sodroski, J. (1994). Probing the structure of the human immunodeficiency virus surface glycoprotein gp120 with a panel of monoclonal antibodies. *J. Virol.* **68**, 469-484.
- Mooslehner, K., Karls, U., and Harbers, K. (1990). Retroviral integration sites in transgenic Mov mice frequently map in the vicinity of transcribed DNA regions. *J. Virol.* **64**, 3056-3058.
- Morris, A., Marsden, M., Halcrow, K., Hughes, E. S., Brettle, R. P., Bell, J. E., and Simmonds, P. (1999). Mosaic structure of the human immunodeficiency virus type 1 genome infecting lymphoid

- cells and the brain: Evidence for frequent in vivo recombination events in the evolution of regional populations. *J Virol* **73**, 8720-8731.
- Moulard, M., Lortat-Jacob, H., Mondor, I., Roca, G., Wyatt, R., Sodroski, J., Zhao, L., Olson, W., Kwong, P. D., and Sattentau, Q. J. (2000). Selective interactions of polyanions with basic surfaces on human immunodeficiency virus type 1 gp120. *J. Virol.* **74**, 1948-1960.
- Moutouh, L., Corbeil, J., and Richman, D. D. (1996). Recombination leads to the rapid emergence of HIV-1 dually resistant mutants under selective drug pressure. *Proc.Natl.Acad.Sci.USA* **93**, 6106-6111.
- Muesing, M. A., Smith, D. H., Cabradilla, C. D., Benton, C. V., Lasky, L. A., and Capon, D. J. (1985). Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. *Nature* **313**, 450-458.
- Mukhtar, M., Harley, S., Chen, P., Bouhamdan, M., Patel, C., Acheampong, E., and Pomerantz, R. J. (2002). Primary isolated human brain microvascular endothelial cells express diverse HIV/SIV-associated chemokine coreceptors and DC-SIGN and L-SIGN. *Virology* **297**, 78-88.
- Murakami, T. and Freed, E. O. (2000). Genetic evidence for an interaction between human immunodeficiency virus type 1 matrix and alpha-helix 2 of the gp41 cytoplasmic tail. *J.Virol.* **74**, 3548-3554.
- Muriaux, D., Derocquigny, H., Roques, B. P., and Paoletti, J. (1996). NCp7 activates HIV-1 (Lai) RNA dimerization by converting a transient loop-loop complex into a stable dimer. *J Biol Chem* **271**, 33686-33692.
- Murray, V. (1989). Improved double-stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Res.* **17**, 8889.
- Myers, G., Korber, B., Smith, R. F., Berzofsky, J. A., and Pavlakis, G. N. (1992). In "Human retroviruses and AIDS." (G. Myers, B. Korber, R. F. Smith, J. Berzofsky, and G. N. Pavlakis, Eds.), Los Alamos, New Mexico: Los Alamos national laboratory.
- Myszka, D. G., Sweet, R. W., Hensley, P., Brigham-Burke, M., Kwong, P. D., Hendrickson, W. A., Wyatt, R., Sodroski, J., and Doyle, M. L. (2000). Energetics of the HIV gp120-CD4 binding reaction. *Proc.Natl.Acad.Sci.U.S.A* **97**, 9026-9031.
- Nabel, G. J. (2001). Challenges and opportunities for development of an AIDS vaccine. *Nature* **410**, 1002-1007.
- Najera, I., Holguin, A., Quinonesmateu, M. E., Munozfernandez, M. A., Najera, R., Lopezgalindez, 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**, 23-31.
- Nakata, K., Weiden, M., Harkin, T., Ho, D., and Rom, W. N. (1995). Low copy number and limited variability of proviral DNA in alveolar macrophages from HIV-1-infected patients: evidence for genetic differences in HIV-1 between lung and blood macrophage populations. *Mol.Med.* **1**, 744-757.
- Nath, A., Psooy, K., Martin, C., Knudsen, B., Magnuson, D. S. K., Haughey, N., and Geiger, J. D. (1996). Identification of a human immunodeficiency virus type 1 tat epitope that is neuroexcitatory and neurotoxic. *J.Virol.* **70**, 1475-1480.
- Navia, B. A. and Price, R. W. (1987). The acquired immunodeficiency syndrome dementia complex as the presenting or sole manifestation of human immunodeficiency virus infection. *Arch.Neurol.* **44**, 65-69.

- Negrone, M. and Buc, H. (1999). Recombination during reverse transcription: an evaluation of the role of the nucleocapsid protein. *J.Mol.Biol.* **286**, 15-31.
- Neumann, M., Felber, B. K., Kleinschmidt, A., Froese, B., Erfle, V., Pavlakis, G. N., and Brackwerner, R. (1995). Restriction of human immunodeficiency virus type 1 production in a human astrocytoma cell line is associated with a cellular block in rev function. *J.Virol.* **69**, 2159-2167.
- O'Brien, S. J. and Moore, J. P. (2000). The effect of genetic variation in chemokines and their receptors on HIV transmission and progression to AIDS. *Immunol.Rev.* **177**, 99-111.
- O'Brien, W. A., Koyanagi, Y., Namazie, A., Zhao, J. Q., Diagne, A., Idler, K., Zack, J. A., and Chen, I. S. (1990). HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120 outside the CD4-binding domain. *Nature.* **348**, 69-73.
- Odelberg, S. J., Weiss, R. B., Hata, A., and White, R. (1995). Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase I. *Nucleic Acids Res.* **23**, 2049-2057.
- Ohagen, A., Ghosh, S., He, J., Huang, K., Chen, Y., Yuan, M., Osathanondh, R., Gartner, S., Shi, B., Shaw, G., and Gabuzda, D. (1999). Apoptosis induced by infection of primary brain cultures with diverse human immunodeficiency virus type 1 isolates: evidence for a role of the envelope. *J.Virol.* **73**, 897-906.
- Olsen, D. B. and Eckstein, F. (1989). Incomplete primer extension during in vitro DNA amplification catalyzed by Taq polymerase; exploitation for DNA sequencing. *Nucleic Acids Res.* **17**, 9613-9620.
- Ono, A. and Freed, E. O. (1999). Binding of human immunodeficiency virus type 1 Gag to membrane: role of the matrix amino terminus. *J.Virol.* **73**, 4136-4144.
- Ono, A., Huang, M. J., and Freed, E. O. (1997). Characterization of human immunodeficiency virus type 1 matrix revertants: Effects on virus assembly, Gag processing, and env incorporation into virions. *J.Virol* **71**, 4409-4418.
- Ono, A., Orenstein, J. M., and Freed, E. O. (2000). Role of the Gag matrix domain in targeting human immunodeficiency virus type 1 assembly. *J.Virol.* **74**, 2855-2866.
- Page, R. D. M. and Holmes, E. C. (1998a). Inferring Molecular Phylogeny. In "Molecular Evolution a Phylogenetic Approach" pp. 172-227. Blackwell Science Ltd, Oxford.
- Page, R. D. M. and Holmes, E. C. (1998b). Models of Molecular Evolution. In "Molecular Evolution a Phylogenetic Approach" pp. 228-279. Blackwell Science Ltd, Oxford.
- Palaniappan, C., Wisniewski, M., Wu, W. M., Fay, P. J., and Bambara, R. A. (1996). Misincorporation by HIV-1 reverse transcriptase promotes recombination via strand transfer synthesis. *Journal of Biological Chemistry* **271**, 22331-22338.
- Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le, T., I, Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., and Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739-745.
- Palker, T. J., Clark, M. E., Langlois, A. J., Matthews, T. J., Weinhold, K. J., Randall, R. R., Bolognesi, D. P., and Haynes, B. F. (1988). Type-specific neutralization of the human immunodeficiency virus with antibodies to env-encoded synthetic peptides. *Journal Name Lost*.
- Palmer, D. L., Hjelle, B. L., Wiley, C. A., Allen, S., Wachsman, W., Mills, R. G., Davis, L. E., and Merlin, T. L. (1994). HIV-1 infection despite immediate combination antiviral therapy after infusion of contaminated white cells. *Am.J.Med.* **97**, 289-295.

- Pandori, M. W., Fitch, N. J. S., Craig, H. M., Richman, D. D., Spina, C. A., and Guatelli, J. C. (1996). Producer-cell modification of human immunodeficiency virus type 1: Nef is a virion protein. *Journal of Virology* **70**, 4283-4290.
- Pang, S., Schlesinger, Y., Daar, E. S., Moudgil, T., Ho, D. D., and Chen, I. S. (1992). Rapid generation of sequence variation during primary HIV-1 infection. *AIDS* **6**, 453-460.
- Pang, S., Vinters, H. V., Akashi, T., O'Brien, W. A., and Chen, I. S. (1991). HIV-1 env sequence variation in brain tissue of patients with AIDS-related neurologic disease. *J.Acquir.Immune.Defic.Syndr.* **4**, 1082-1092.
- Panganiban, A. T. and Fiore, D. (1988). Ordered interstrand and intrastrand DNA transfer during reverse transcription. *Science* **241**, 1064-1069.
- Pantaleo, G., Graziosi, C., Demarest, J. F., Butini, L., Montroni, M., Fox, C. H., Orenstein, J. M., Kotler, D. P., and Fauci, A. S. (1993). HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease. *Nature* **362**, 355-358.
- Pantaleo, G., Menzo, S., Vaccarezza, M., Graziosi, C., Cohen, O. J., Demarest, J. F., Montefiori, D., Orenstein, J. M., Fox, C., Schrager, L. K., Margolick, J. B., Buchbinder, S., Giorgi, J. V., and Fauci, A. S. (1995). Studies in subjects with long term nonprogressive human immunodeficiency virus infection. *N.Engl.J.Med.* **332**, 209-216.
- Paquette, J. S., Fortin, J. F., Blanchard, L., and Tremblay, M. J. (1998). Level of ICAM-1 surface expression on virus producer cells influences both the amount of virion-bound host ICAM-1 and human immunodeficiency virus type 1 infectivity. *J Virol* **72**, 9329-9336.
- Patel, M., Yanagishita, M., Roderiquez, G., Bouhabib, D. C., Oravec, T., Hascall, V. C., and Norcross, M. A. (1993). Cell-surface heparan sulfate proteoglycan mediates HIV-1 infection of t-cell lines. *AIDS Res.Hum.Retroviruses* **9**, 167-174.
- Patel, P. H. and Preston, B. D. (1994). Marked infidelity of human immunodeficiency virus type 1 reverse transcriptase at RNA and DNA template ends. *Proc.Natl.Acad.Sci.USA* **91**, 549-553.
- Pathak, V. K. and Temin, H. M. (1990a). Broad spectrum of *in vivo* forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: deletions and deletions with insertions. *Proc.Natl.Acad.Sci.U.S.A* **87** , 6024-6028.
- Pathak, V. K. and Temin, H. M. (1990b). Broad spectrum of *in vivo* forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. *Proc.Natl.Acad.Sci.USA.* **87**, 6019-6023.
- Paxton, W., Connor, R. I., and Landau, N. R. (1993). Incorporation of vpr into human immunodeficiency virus type-1 virions - requirement for the p6 region of gag and mutational analysis. *J.Virol.* **67**, 7229-7237.
- Peckham, C. and Gibb, D. (1995). Current concepts - mother-to-child transmission of the human immunodeficiency virus. *N.Engl.J.Med.* **333**, 298-302.
- Peden, K. W. (1992). Instability of HIV sequences in high copy number plasmids. *J.Acquir.Immune.Defic.Syndr.* **5**, 313-315.
- Peeters, M., Gueye, A., Mboup, S., Bibollettruche, F., Ekaza, E., Mulanga, C., Ouerdrago, R., Gandji, R., MPele, P., Dibanga, G., Koumare, B., Saidou, M., Esuwilliams, E., Lombart, J. P., Badombena, W., Luo, N., Vandenhaesevelde, M., and Delaporte, E. (1997). Geographical distribution of HIV-1 group O viruses in Africa. *AIDS* **11**, 493-498.

- Peeters, M., Liegeois, F., Torimiro, N., Bourgeois, A., Mpoudi, E., Vergne, L., Saman, E., Delaporte, E., and Saragosti, S. (1999). Characterization of a highly replicative intergroup M/O human immunodeficiency virus type 1 recombinant isolated from a Cameroonian patient. *J Virol* **73**, 7368-7375.
- Peeters, M. and Sharp, P. M. (2000). Genetic diversity of HIV-1: the moving target. *AIDS* **14 Suppl 3**, S129-S140.
- Perelson, A. S., Essunger, P., Cao, Y. Z., Vesanen, M., Hurley, A., Saksela, K., Markowitz, M., and Ho, D. D. (1997). Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* **387**, 188-191.
- Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M., and Ho, D. D. (1996). HIV-1 dynamics in vivo: Virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**, 1582-1586.
- Petito, C. K., Chen, H., Mastri, A. R., Torres-Munoz, J., Roberts, B., and Wood, C. (1999). HIV infection of choroid plexus in AIDS and asymptomatic HIV- infected patients suggests that the choroid plexus may be a reservoir of productive infection. *J.Neurovirol.* **5**, 670-7.
- Phillips, A. N. (1996). Reduction of HIV concentration during acute infection: Independence from a specific immune response. *Science* **271**, 497-499.
- Phillips, R. E., Rowland Jones, S., Nixon, D. F., Gotch, F. M., Edwards, J. P., Ogunlesi, A. O., Elvin, J. G., Rothbard, J. A., Bangham, C. R., and Rizza, C. R. (1991). Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature* **354**, 453-459.
- Philpott, S., Burger, H., Charbonneau, T., Grimson, R., Vermund, S. H., Visosky, A., Nachman, S., Kovacs, A., Tropper, P., Frey, H., and Weiser, B. (1999). CCR5 genotype and resistance to vertical transmission of HIV-1. *J Acq Immun Defic Synd Hum R* **21**, 189-193.
- Picard, L., Simmons, G., Power, C. A., Meyer, A., Weiss, R. A., and Clapham, P. R. (1997). Multiple extracellular domains of CCR-5 contribute to human immunodeficiency virus type 1 entry and fusion. *J Virol* **71**, 5003-5011.
- Pignet, V., Gu, F., Foti, M., Demareux, N., Gruenberg, J., Carpentier, J. L., and Trono, D. (1999). Nef-induced CD4 degradation: A diacidic-based motif in Nef functions as a lysosomal targeting signal through the binding of beta-COP in endosomes. *Cell* **97**, 63-73.
- Piller, S. C., Jans, P., Gage, P. W., and Jans, D. A. (1998). Extracellular HIV-1 virus protein R causes a large inward current and cell death in cultured hippocampal neurons: Implications for AIDS pathology. *Proc Natl Acad Sci USA* **95**, 4595-4600.
- Ping, L. H., Cohen, M. S., Hoffman, I., Vernazza, P., Seillier-Moiseiwitsch, F., Chakraborty, H., Kazembe, P., Zimba, D., Maida, M., Fiscus, S. A., Eron, J. J., Swanstrom, R., and Nelson, J. A. (2000). Effects of genital tract inflammation on human immunodeficiency virus type 1 V3 populations in blood and semen. *J.Virol.* **74**, 8946-8952.
- Platt, E. J., Madani, N., Kozak, S. L., and Kabat, D. (1997). Infectious properties of human immunodeficiency virus type 1 mutants with distinct affinities for the CD4 receptor. *J Virol* **71**, 883-890.
- Platt, E. J., Wehrly, K., Kuhmann, S. E., Chesebro, B., and Kabat, D. (1998). Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. *J Virol* **72**, 2855-2864.

- Pohlmann, S., Baribaud, F., and Doms, R. W. (2001a). DC-SIGN and DC-SIGNR: helping hands for HIV. *Trends Immunol.* **22**, 643-646.
- Pohlmann, S., Soilleux, E. J., Baribaud, F., Leslie, G. J., Morris, L. S., Trowsdale, J., Lee, B., Coleman, N., and Doms, R. W. (2001b). DC-SIGNR, a DC-SIGN homologue expressed in endothelial cells, binds to human and simian immunodeficiency viruses and activates infection in trans. *Proc.Natl.Acad.Sci.U.S.A* **98**, 2670-2675.
- Pomerantz, R. J., Trono, D., Feinberg, M. B., and Baltimore, D. (1990). Cells nonproductively infected with HIV-1 exhibit an aberrant pattern of viral RNA expression: a molecular model for latency. *Cell* **61**, 1271-1276.
- Poon, B., Grovitferbas, K., Stewart, S. A., and Chen, I. S. Y. (1998). Cell cycle arrest by Vpr in HIV-1 virions and insensitivity to antiretroviral agents. *Science* **281**, 266-269.
- Poon, D. T. K., Wu, J., and Aldovini, A. (1996). Charged amino acid residues of human immunodeficiency virus type 1 nucleocapsid p7 protein involved in RNA packaging and infectivity. *J.Virol.* **70**, 6607-6616.
- Popov, S., Rexach, M., Ratner, L., Blobel, G., and Bukrinsky, M. (1998a). Viral protein R regulates docking of the HIV-1 preintegration complex to the nuclear pore complex. *J Biol Chem* **273**, 13347-13352.
- Popov, S., Rexach, M., Zybarth, G., Reiling, N., Lee, W. A., Ratner, L., Lane, C. M., Moore, M. S., Blobel, G., and Bukrinsky, M. (1998b). Viral protein R regulates nuclear import of the HIV-1 preintegration complex. *EMBO J* **17**, 909-917.
- Poss, M., Rodrigo, A. G., Gosink, J. J., Learn, G. H., Panteleeff, D. D., Martin, H. L., Bwayo, J., Kreiss, J. K., and Overbaugh, J. (1998). Evolution of envelope sequences from the genital tract and peripheral blood of women infected with clade A human immunodeficiency virus type 1. *J Virol* **72**, 8240-8251.
- Power, C., Mearthur, J. C., Johnson, R. T., Griffin, D. E., Glass, J. D., Perryman, S., and Chesebro, B. (1994). Demented and nondemented patients with AIDS differ in brain-derived human immunodeficiency virus type 1 envelope sequences. *J.Virol.* **68**, 4643-4649.
- Power, C., Mearthur, J. C., Nath, A., Wehrly, K., Mayne, M., Nishio, J., Langelier, T., Johnson, R. T., and Chesebro, B. (1998). Neuronal death induced by brain-derived human immunodeficiency virus type 1 envelope genes differs between demented and nondemented AIDS patients. *J Virol* **72**, 9045-9053.
- Preston, B. D. and Dougherty, J. P. (1996). Mechanisms of retroviral mutation. *Trends Microbiol.* **4**, 16-21.
- Preston, B. D., Poiesz, B. J., and Loeb, L. A. (1988). Fidelity of HIV-1 Reverse Transcriptase. *Science* **242**, 1168-1171.
- Price, D. A., Goulder, P. J. R., Klenerman, P., Sewell, A. K., Easterbrook, P. J., Troop, M., Bangham, C. R. M., and Phillips, R. E. (1997). Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proc Natl Acad Sci USA* **94**, 1890-1895.
- Price, R. W. (1996). Neurological complications of HIV infection. *Lancet* **348**, 445-452.
- Price, R. W. (2000). The two faces of HIV infection of cerebrospinal fluid. *Trends Microbiol.* **8**, 387-391.

- Price, R. W., Brew, B., Sidtis, J., Rosenblum, M., Scheck, A. C., and Cleary, P. (1988). The Brain in AIDS: Central Nervous System HIV-1 and AIDS Dementia Complex. *Science* **239**, 586-592.
- Pryciak, P. M., Sil, A., and Varmus, H. E. (1992). Retroviral integration into minichromosomes in vitro. *EMBO J.* **11**, 291-303.
- Puffer, B. A., Sharron, M., Coughlan, C. M., Baribaud, F., McManus, C. M., Lee, B., David, J., Price, K., Horuk, R., Tsang, M., and Doms, R. W. (2000). Expression and coreceptor function of APJ for primate immunodeficiency viruses. *Virology* **276**, 435-444.
- Pumarola Sune, T., Navia, B. A., Cordon Cardo, C., Cho, E. S., and Price, R. W. (1987). HIV antigen in the brains of patients with the AIDS dementia complex. *Ann.Neurol.* **21**, 490-496.
- Purcell, D. F. J. and Martin, M. A. (1993). Alternative splicing of human immunodeficiency virus type-1 messenger RNA modulates viral protein expression, replication, and infectivity. *J.Virol.* **67**, 6365-6378.
- Quinn, T. C. (1996). Global burden of the HIV pandemic. *Lancet* **348**, 99-106.
- Ranki, A., Nyberg, M., Ovod, V., Haltia, M., Elovaara, I., Raininko, R., Haapasalo, H., and Krohn, K. (1995). Abundant expression of HIV nef and rev proteins in brain astrocytes in vivo is associated with dementia. *AIDS* **9**, 1001-1008.
- Ratner, L., Gallo, R. C., and Wong Staal, F. (1985). HTLV-III, LAV, ARV are variants of same AIDS virus. *Nature* **313**, 636-637.
- Rayfield, M., De Cock, K., Heyward, W., Goldstein, L., Krebs, J., Kwok, S., Lee, S., McCormick, J., Moreau, J. M., Odehouri, K., and . (1988). Mixed human immunodeficiency virus (HIV) infection in an individual: demonstration of both HIV type 1 and type 2 proviral sequences by using polymerase chain reaction. *Journal of Infectious Diseases* **158**, 1170-1176.
- Reddy, R. T., Achim, C. L., Sirko, D. A., Tehranchi, S., Kraus, F. G., Wongstaal, F., Wiley, C. A., Grant, I., Atkinson, J. H., Kelly, M., Chandler, J. L., Wallace, M. R., McCutchan, J. A., Spector, S. A., Thal, L., Heaton, R. K., Hesselink, J., Jernigan, T., Masliah, E., Abramson, I., Butters, N., Patterson, T., Zisook, S., and Jeste, D. (1996). Sequence analysis of the V3 loop in brain and spleen of patients with HIV encephalitis. *AIDS Res.Hum.Retroviruses* **12**, 477-482.
- Reeves, J. D., McKnight, A., Potempa, S., Simmons, G., Gray, P. W., Power, C. A., Wells, T., Weiss, R. A., and Talbot, S. J. (1997). CD4-independent infection by HIV-2 (ROD/B): Use of the 7-transmembrane receptors CXCR-4, CCR-3, and V28 for entry. *Virology* **231**, 130-134.
- Rhee, S. S. and Marsh, J. W. (1994). Human immunodeficiency virus type 1 nef-induced down-modulation of CD4 is due to rapid internalization and degradation of surface CD4. *J.Virol.* **68**, 5156-5163.
- Rhodes, D. I., Ashton, L., Solomon, A., Carr, A., Cooper, D., Kaldor, J., and Deacon, N. (2000). Characterization of three nef-defective human immunodeficiency virus type 1 strains associated with long-term nonprogression. Australian Long-Term Nonprogressor Study Group. *J.Virol.* **74**, 10581-10588.
- Richman, D. D. (2001). HIV chemotherapy. *Nature* **410**, 995-1001.
- Richman, D. D., Havlir, D., Corbeil, J., Looney, D., Ignacio, C., Spector, S. A., Sullivan, J., Cheeseman, S., Barringer, K., Pauletti, D., Shih, C. K., Myers, M., and Griffin, J. (1994). Nevirapine resistance mutations of human immunodeficiency virus type 1 selected during therapy. *J.Virol.* **68**, 1660-1666.

- Rimsky, L. T., Shugars, D. C., and Matthews, T. J. (1998). Determinants of human immunodeficiency virus type 1 resistance to gp41-derived inhibitory peptide. *J Virol* **72**, 986-993.
- Rist, M. J. and Marino, J. P. (2002). Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1. *Biochemistry* **41**, 14762-14770.
- Rizzuto, C. D., Wyatt, R., HernandezRamos, N., Sun, Y., Kwong, P. D., Hendrickson, W. A., and Sodroski, J. (1998). A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science* **280**, 1949-1953.
- Roberts, J. D., Bebenek, K., and Kunkel, T. A. (1988). The accuracy of reverse transcriptase from HIV-1. *Science* **242**, 1171-1173.
- Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., Gao, F., Hahn, B. H., Kalish, M. L., Kuiken, C., Learn, G. H., Leitner, T., Mccutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P. M., Wolinsky, S., and Korber, B. (2000). HIV-1 nomenclature proposal. *Science* **288**, 55-56.
- Robertson, D. L., Sharp, P. M., McCutchan, F. E., and Hahn, B. H. (1995). Recombination of HIV-1. *Nature* **374**, 124-126.
- Roderiquez, G., Oravec, T., Yanagishita, M., Bouhabib, D. C., Mostowski, H., and Norcross, M. A. (1995). Mediation of human immunodeficiency virus type 1 binding by interaction of cell surface heparan sulfate proteoglycans with the v3 region of envelope gp120-gp41. *J Virol* **69**, 2233-2239.
- Rodrigo, A. G., Hanley, E. W., Goracke, P. C., and Learn, G. H. (2001). Sampling and processing HIV molecular sequences: a computational evolutionary biologist's perspective. In "Computational and evolutionary analysis of HIV molecular sequences" (A. G. Rodrigo and G. H. Learn, Eds.), pp. 1-17. Kluwer Academic Publishers, Norwell, Massachusetts, USA.
- Rodrigo, A. G., Goracke, P. C., Rowhanian, K., and Mullins, J. I. (1997). Quantitation of target molecules from polymerase chain reaction- based limiting dilution assays. *AIDS Res Hum Retroviruses* **13**, 737-742.
- RodriguezFrade, J. M., VilaCoro, A. J., Martin, A., Nieto, M., SanchezMadrid, F., Proudfoot, A. E. I., Wells, T. N. C., Martinez, C., and Mellado, M. (1999). Similarities and differences in RANTES- and (AOP)-RANTES- triggered signals: Implications for chemotaxis. *J Cell Biol* **144**, 755-765.
- Rodriguezrodriguez, L., Tsuchihashi, Z., Fuentes, G. M., Bambara, R. A., and Fay, P. J. (1995). Influence of human immunodeficiency virus nucleocapsid protein on synthesis and strand transfer by the reverse transcriptase in vitro. *Journal of Biological Chemistry* **270**, 15005-15011.
- Rollins, B. J. (1997). Chemokines. *Blood* **90**, 909-928.
- Rose, J. K. and Shafferman, A. (1981). Conditional expression of the vesicular stomatitis virus glycoprotein gene in Escherichia coli. *Proc.Natl.Acad.Sci.U.S.A* **78**, 6670-6674.
- Ross, H. L., Gartner, S., Mearthur, J. C., Corboy, J. R., McAllister, J. J., Millhouse, S., and Wigdahl, B. (2001). HIV-1 LTR C/EBP binding site sequence configurations preferentially encountered in brain lead to enhanced C/EBP factor binding and increased LTR-specific activity. *J.Neurovirol.* **7**, 235-249.
- Ross, T. M. and Cullen, B. R. (1998). The ability of HIV type 1 to use CCR-3 as a coreceptor is controlled by envelope V1/V2 sequences acting in conjunction with a CCR-5 tropic V3 loop. *Proc Natl Acad Sci USA* **95**, 7682-7686.

- Roth, M. J., Schwartzberg, P. L., and Goff, S. P. (1989). Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence. *Cell* **58**, 47-54.
- Rouzine, I. M. and Coffin, J. M. (1999). Linkage disequilibrium test implies a large effective population number for HIV in vivo. *Proc.Natl.Acad.Sci.U.S.A* **96**, 10758-10763.
- Rouzine, I. M., Rodrigo, A., and Coffin, J. M. (2001). Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology. *Microbiol.Mol.Biol.Rev.* **65**, 151-185.
- Rucker, J. and Doms, R. W. (1998). Chemokine receptors as HIV coreceptors: Implications and interactions. *AIDS Res Hum Retroviruses* **14**, S241-S246.
- Rucker, J., Edinger, A. L., Sharron, M., Samson, M., Lee, B., Berson, J. F., Yi, Y. J., Margulies, B., Collman, R. G., Doranz, B. J., Parmentier, M., and Doms, R. W. (1997). Utilization of chemokine receptors, orphan receptors, and herpesvirus- encoded receptors by diverse human and simian immunodeficiency viruses. *J Virol* **71**, 8999-9007.
- Rusche, J. R., Javaherian, K., McDanal, C., Petro, J., Lynn, D. L., Grimaila, R., Langlois, A., Gallo, R. C., Arthur, L. O., Fischinger, P. J., Bolognesi, D. P., Putney, S. D., and Matthews, T. J. (1988). 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.U.S.A.* **85**, 3198-3202.
- Rzhetsky, A. and Nei, M. (1992). A simple method for estimating and testing minimum-evolution trees. *Molecular Biology and Evolution* **9**, 945-967.
- Saag, M. S., Hahn, B. H., Gibbons, J., Li, Y., Parks, E. S., Parks, W. P., and Shaw, G. M. (1988). Extensive variation of human immunodeficiency virus type-1 in vivo. *Nature* **334**, 440-444.
- Sabatier, J. M., Vives, E., Mabrouk, K., Benjouad, A., Rochat, H., Duval, A., Hue, B., and Baharoui, E. (1991). Evidence for neurotoxic activity of *tat* from human immunodeficiency virus type-1. *J.Virol.* **65**, 961-967.
- Sabino, E. C., Shpaer, E. G., Morgado, M. G., Korber, B. T. M., Diaz, R. S., Bongertz, V., Cavalcante, S., Galvaocastro, B., Mullins, J. I., and Mayer, A. (1994). Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and Fin two epidemiologically linked individuals from Brazil. *J.Virol.* **68**, 6340-6346.
- Sacktor, N., Lyles, R. H., Skolasky, R., Kleeberger, C., Selnes, O. A., Miller, E. N., Becker, J. T., Cohen, B., and McArthur, J. C. (2001). HIV-associated neurologic disease incidence changes:: Multicenter AIDS Cohort Study, 1990-1998. *Neurology* **56**, 257-260.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science.* **239**, 487-491.
- Saitou, N. and Imanishi, M. (1989). Relative efficiencies of the Fitch-Margolish, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbor-joining methods of phylogenetic tree reconstruction in obtaining the correct tree. *Molecular Biology and Evolution* **6**, 514-525.
- Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol.Biol.Evol.* **4**, 406-425.
- Salminen, M. O., Carr, J. K., Burke, D. S., and McCutchan, F. E. (1995). Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res.Hum.Retroviruses* **11**, 1423-1425.

- Salzwedel, K. and Berger, E. A. (2000). Cooperative subunit interactions within the oligomeric envelope glycoprotein of HIV-1: functional complementation of specific defects in gp120 and gp41. *Proc.Natl.Acad.Sci.U.S.A* **97**, 12794-12799.
- Samson, M., Edinger, A. L., Stordeur, P., Rucker, J., Verhasselt, V., Sharron, M., Govaerts, C., Mollereau, C., Vassart, G., Doms, R. W., and Parmentier, M. (1998). ChemR23, a putative chemoattractant receptor, is expressed in monocyte-derived dendritic cells and macrophages and is a coreceptor for SIV and some primary HIV-1 strains. *Eur.J.Immunol.* **28**, 1689-1700.
- Samuel, K. P., Seth, A., Zweig, M., Showalter, S. D., and Papas, T. S. (1988). Bacterial expression and characterization of nine polypeptides encoded by segments of the envelope gene of human immunodeficiency virus. *Gene* **64**, 121-134.
- Sanchez Pescador, R., Power, M. D., Barr, P. J., Steimer, K. S., Stempien, M. M., Brown Shimer, S. L., Gee, W. W., Renard, A., Randolph, A., and Levy, J. A. (1985). Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). *Science* **227**, 484-492.
- Sanchezpalomino, S., Rojas, J. M., Martinez, M. A., Fenyo, E. M., Najera, R., Domingo, E., and Lopezgalindez, C. (1993). Dilute passage promotes expression of genetic and phenotypic variants of human immunodeficiency virus type-1 in cell culture. *J.Virol.* **67**, 2938-2943.
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc.Natl.Acad.Sci.U.S.A* **74**, 5463-5467.
- Sattentau, Q. J. and Moore, J. P. (1991). Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. *J.Exp.Med.* **174**, 407-415.
- Sattentau, Q. J., Moore, J. P., Vignaux, F., Traincard, F., and Poignard, P. (1993). Conformational changes induced in the envelope glycoproteins of the human and simian immunodeficiency viruses by soluble receptor binding. *J.Virol.* **67**, 7383-7393.
- Sayasith, K., Sauve, G., and Yelle, J. (2000). Characterization of mutant HIV-1 integrase carrying amino acid changes in the catalytic domain. *Mol.Cells* **10**, 525-532.
- Scarlatti, G., Tresoldi, E., Bjorndal, A., Fredriksson, R., Colognesi, C., Deng, H. K., Malnati, M. S., Plebani, A., Siccardi, A. G., Littman, D. R., Fenyo, E. M., and Lusso, P. (1997). In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression. *Nature Med* **3**, 1259-1265.
- Schacker, T., Collier, A. C., Hughes, J., Shea, T., and Corey, L. (1996). Clinical and epidemiologic features of primary HIV infection. *Ann.Intern.Med* **125**, 257.
- Scherdin, V., Rhodes, K., and Breindl, M. (1990). Transcriptionally active genome regions are preferred targets for retrovirus integration. *J.Virol.* **64**, 907-912.
- Schim van der Loeff MF and Aaby, P. (1999). Towards a better understanding of the epidemiology of HIV-2. *AIDS* **13 Suppl A**, S69-S84.
- Schmidtayerova, H., Alfano, M., Nuovo, G., and Bukrinsky, M. (1998). Human immunodeficiency virus type 1 T-lymphotropic strains enter macrophages via a CD4- and CXCR4-Mediated pathway: Replication is restricted at a post-entry level. *J Virol* **72**, 4633-4642.
- Schubert, U., Ferrer-Montiel, A. V., Oblatt-Montal, M., Henklein, P., Strebel, K., and Montal, M. (1996). 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.* **398**, 12-18.

- Schubert, U., Henklein, P., Boldyreff, B., Wingender, E., Strebel, K., and Porstmann, T. (1994). The human immunodeficiency virus type 1 encoded Vpu protein is phosphorylated by casein kinase-2 (CK-2) at positions Ser52 and Ser56 within a predicted alpha-helix-turn-alpha-helix-motif. *J.Mol.Biol.* **236**, 16-25.
- Schuitemaker, H., Koot, M., Kootstra, N., Dercksen, M. W., Goede, R. T. D., Steenwijk, R., Lange, J. M. A., Schatttenkerk, J. K., Miedema, F., and Tersmette, M. (1992). Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytotropic to T cell-tropic virus populations. *J.Virol* **66**, 1354-1360.
- Schuler, W., Wecker, K., de Rocquigny, H., Baudat, Y., Sire, J., and Roques, B. P. (1999). NMR structure of the (52-96) C-terminal domain of the HIV-1 regulatory protein Vpr: molecular insights into its biological functions. *J.Mol.Biol.* **285**, 2105-2117.
- Schuurman, R., Nijhuis, M., Vanleeuwen, R., Schipper, P., Dejong, D., Collis, P., Danner, S. A., Mulder, J., Loveday, C., Christopherson, C., Kwok, S., Sninsky, J., and Boucher, C. A. B. (1995). Rapid changes in human immunodeficiency virus type 1 RNA load and appearance of drug-resistant virus populations in persons treated with lamivudine (3TC). *Journal of Infectious Diseases* **171**, 1411-1419.
- Schwartz, S., Felber, B. K., Benko, D. M., Fenyo, E. M., and Pavlakis, G. N. (1990). Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. *J.Virol.* **64**, 2519-2529.
- Schwartz, S., Felber, B. K., Fenyo, E. M., and Pavlakis, G. N. (1989). Rapidly and slowly replicating human immunodeficiency virus type 1 isolates can be distinguished according to target-cell tropism in T-cell and monocyte cell lines. *Proc.Natl.Acad.Sci.U.S.A.* **86**, 7200-7203.
- Seddiki, N., BenYounesChennoufi, A., Benjouad, A., Saffar, L., Baumann, N., Gluckman, J. C., and Gattegno, L. (1996). Membrane glycolipids and human immunodeficiency virus infection of primary macrophages. *AIDS Res.Hum.Retroviruses* **12**, 695-703.
- Seddiki, N., Ramdani, A., Saffar, L., Portoukalian, J., Gluckman, J. C., and Gattegno, L. (1994). A monoclonal antibody directed to sulfatide inhibits the binding of human immunodeficiency virus type 1 (HIV-1) envelope glycoprotein to macrophages but not their infection by the virus. *Biochim.Biophys.Acta* **1225**, 289-296.
- Selig, L., Pages, J. C., Tanchou, V., Preveral, S., BerliozTorrent, C., Liu, L. X., Erdtmann, L., Darlix, J. L., Benarous, R., and Benichou, S. (1999). Interaction with the p6 domain of the Gag precursor mediates incorporation into virions of Vpr and Vpx proteins from primate lentiviruses. *J.Virol* **73**, 592-600.
- Shafikhani, S. (2002). Factors affecting PCR-mediated recombination. *Environ.Microbiol.* **4**, 482-486.
- Shahabuddin, M., Bentsman, G., Volsky, B., Rodriguez, I., and Volsky, D. J. (1996). A mechanism of restricted human immunodeficiency virus type 1 expression in human glial cells. *J.Virol.* **70**, 7992-8002.
- Shankarappa, R., Gupta, P., Learn, G. H., Rodrigo, A. G., Rinaldo, C. R., Gorry, M. C., Mullins, J. I., Nara, P. L., and Ehrlich, G. D. (1998). Evolution of human immunodeficiency virus type 1 envelope sequences in infected individuals with differing disease progression profiles. *Virology* **241**, 251-259.
- Shapshak, P., Segal, D. M., Crandall, K. A., Fujimura, R. K., Zhang, B. T., Xin, K. Q., Okuda, K., Petit, C. K., Eisdorfer, C., and Goodkin, K. (1999). Independent evolution of HIV type 1 in different brain regions. *AIDS Res.Hum.Retroviruses* **15**, 811-20.

- Sharp, P. M., Bailes, E., Robertson, D. L., Gao, F., and Hahn, B. H. (1999). Origins and evolution of AIDS viruses. *Biol Bull* **196**, 338-342.
- Sharp, P. M., Bailes, E., Stevenson, M., Emerman, M., and Hahn, B. H. (1996). Gene acquisition in HIV and SIV. *Nature* **383**, 586-587.
- Sharp, P. M., Robertson, D. L., Feng, G., and Hahn, B. H. (1994). Origins and diversity of human immunodeficiency viruses. *AIDS* **8**, S27-S42.
- Sharron, M., Pohlmann, S., Price, K., Lolis, E., Tsang, M., Kirchhoff, F., Doms, R. W., and Lee, B. (2000). Expression and coreceptor activity of STRL33/Bonzo on primary peripheral blood lymphocytes. *Blood* **96**, 41-49.
- Shaw, G. M., Hahn, B. H., Arya, S. K., Groopman, J. E., Gallo, R. C., and Wong Staal, F. (1984). Molecular characterization of human T-cell leukemia (lymphotropic) virus type III in the acquired immune deficiency syndrome. *Science*. **226**, 1165-1171.
- Shaw, G. M., Harper, M. E., Hahn, B. E., Epstein, L. G., Gajdusek, D., Price, R., Navia, B., Petit, C., O'Hara, C., and Groopman, J. (1985). HTLV-III infection in brains of children and adults with AIDS encephalopathy. *Science*. **1**, 177-182.
- Sheehy, A. M., Gaddis, N. C., Choi, J. D., and Malim, M. H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646-650.
- Shi, B., Raina, J., Lorenzo, A., Busciglio, J., and Gabuzda, D. (1998). Neuronal apoptosis induced by HIV-1 Tat protein and TNF-alpha: potentiation of neurotoxicity mediated by oxidative stress and implications for HIV-1 dementia. *J.Neurovirol.* **4**, 281-290.
- Shi, C. and Mellors, J. W. (1997). A recombinant retroviral system for rapid in vivo analysis of human immunodeficiency virus type 1 susceptibility to reverse transcriptase inhibitors. *Antimicrob.Agents Chemother.* **41**, 2781-5.
- Shieh, J. T., Martin, J., Baltuch, G., Malim, M. H., and Gonzalez-Scarano, F. (2000). Determinants of syncytium formation in microglia by human immunodeficiency virus type 1: role of the V1/V2 domains. *J.Virol.* **74**, 693-701.
- Shieh, J. T. C., Albright, A. V., Sharron, M., Gartner, S., Strizki, J., Doms, R. W., and Gonzalezscarano, F. (1998). Chemokine receptor utilization by human immunodeficiency virus type 1 isolates that replicate in microglia. *J Virol* **72**, 4243-4249.
- Shih, C. C., Stoye, J. P., and Coffin, J. M. (1988). Highly preferred targets for retrovirus integration. *Cell* **53**, 531-537.
- Shimizu, N., Soda, Y., Kanbe, K., Liu, H. Y., Mukai, R., Kitamura, T., and Hoshino, H. (2000). A putative G protein-coupled receptor, RDC1, is a novel coreceptor for human and simian immunodeficiency viruses. *J.Virol.* **74**, 619-626.
- Shioda, T., Levy, J. A., and Cheng-Mayer, C. (1992). Small amino acid changes in the V3 hypervariable region of gp120 can affect the T-cell line and macrophage tropism of human immunodeficiency virus type 1. *Proc.Natl.Acad.Sci.USA.* **89**, 9434-9438.
- Shioda, T., Levy, J. A., and Cheng Mayer, C. (1991). Macrophage and T cell-line tropisms of HIV-1 are determined by specific regions of the envelope gp120 gene. *Nature.* **349**, 167-169.
- Sigurdsson, B. (1954). Observations on three slow infections of sheep. Maedi, paratuberculosis, rida, a slow encephalitis of sheep with general remarks on infections which develop slowly and some of their special characteristics. *British Veterinary Journal* **110**, 255-270.

- Simm, M., Shahabuddin, M., Chao, W., Allan, J. S., and Volsky, D. J. (1995). Aberrant gag protein composition of a human immunodeficiency virus type 1 vif mutant: produced in primary lymphocytes. *J.Virol.* **69**, 4582-4586.
- Simmonds, P., Balfe, P., Peutherer, J. F., Ludlam, C. A., Bishop, J. O., and Leigh Brown, A. J. (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.
- Simmonds, P., Zhang, L. Q., McOmish, F., Balfe, P., Ludlam, C. A., and Leigh Brown, A. J. (1991). Discontinuous sequence change of human immunodeficiency virus (HIV) type 1 env sequences in plasma viral and lymphocyte-associated proviral populations in vivo: implications for models of HIV pathogenesis. *J.Virol.* **65**, 6266-6276.
- Simmons, G., Reeves, J. D., McKnight, A., Dejuq, N., Hibbitts, S., Power, C. A., Aarons, E., Schols, D., Declercq, E., Proudfoot, A. E. I., and Clapham, P. R. (1998). CXCR4 as a functional coreceptor for human immunodeficiency virus type 1 infection of primary macrophages. *J Virol* **72**, 8453-8457.
- Simon, F., Mauclore, P., Roques, P., Loussertajaka, I., MullerTrutwin, M. C., Saragosti, S., Georgescourbot, M. C., Barresinoussi, F., and Brunvezinet, F. (1998). Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nature Med* **4**, 1032-1037.
- Sinkora, M., Sun, J., and Butler, J. E. (2000). Antibody repertoire development in fetal and neonatal piglets. V. VDJ gene chimeras resembling gene conversion products are generated at high frequency by PCR in vitro. *Mol.Immunol.* **37**, 1025-1034.
- Siomi, H., Shida, H., Maki, M., and Hatanaka, M. (1990). Effects of a highly basic region of human immunodeficiency virus Tat protein on nucleolar localization. *J.Virol.* **64**, 1803-1807.
- Smith, D. B., Mcallister, J., Casino, C., and Simmonds, P. (1997). Virus 'quasispecies': Making a mountain out of a molehill? *J Gen Virol* **78**, 1511-1519.
- Smith, M. O., Heyes, M. P., and Lackner, A. A. (1995). Early intrathecal events in rhesus macaques (*Macaca mulatta*) infected with pathogenic or nonpathogenic molecular clones of simian immunodeficiency virus. *Lab Invest* **72**, 547-558.
- Snider, W. D., Simpson, D. M., Nielsen, S., Gold, J. W., Metroka, C. E., and Posner, J. B. (1983). Neurological complications of acquired immune deficiency syndrome: analysis of 50 patients. *Ann.Neurol.* **14**, 403-418.
- Soda, Y., Shimizu, N., Jinno, A., Liu, H. Y., Kanbe, K., Kitamura, T., and Hoshino, H. (1999). Establishment of a new system for determination of coreceptor usages of HIV based on the human glioma NP-2 cell line. *Biochem Biophys Res Commun* **258**, 313-321.
- Sodroski, J., Goh, W. C., Rosen, C., Dayton, A., Terwilliger, E., and Haseltine, W. (1986). A second post-transcriptional trans-activator gene required for HTLV-III replication. *Nature.* **321**, 412-417.
- Sommerfelt, M. A. and Asjo, B. (1995). Intercellular adhesion molecule 3, a candidate human immunodeficiency virus type 1 co-receptor on lymphoid and monocytoid cells. *J.Gen.Virol.* **76**, 1345-1352.
- Sonza, S., Maerz, A., Uren, S., Violo, A., Hunter, S., Boyle, W., and Crowe, S. (1995). Susceptibility of human monocytes to HIV type 1 infection in vitro is not dependent on their level of CD4 expression - short communication. *AIDS Res.Hum.Retroviruses* **11**, 769-776.
- SotoRamirez, L. E., Renjifo, B., McLane, M. F., Marlink, R., OHara, C., Sutthent, R., Wasi, C., Vithayasai, P., Vithayasai, V., Apichartpiyakul, C., Auewarakul, P., Cruz, V. P., Chui, D. S.,

- Osathanondh, R., Mayer, K., Lee, T. H., and Essex, M. (1996). HIV-1 Langerhans' cell tropism associated with heterosexual transmission of HIV. *Science* **271**, 1291-1293.
- Speck, R. F., Esser, U., Penn, M. L., Eckstein, D. A., Pulliam, L., Chan, S. Y., and Goldsmith, M. A. (1999). A trans-receptor mechanism for infection of CD4-negative cells by human immunodeficiency virus type 1. *Curr. Biol.* **9**, 547-550.
- Speck, R. F., Wehrly, K., Platt, E. J., Atchison, R. E., Charo, I. F., Kabat, D., Chesebro, B., and Goldsmith, M. A. (1997). Selective employment of chemokine receptors as human immunodeficiency virus type 1 coreceptors determined by individual amino acids within the envelope V3 loop. *J. Virol.* **71**, 7136-7139.
- Srinivasakumar, N., Hammarskjold, M. L., and Rekosh, D. (1995). Characterization of deletion mutations in the capsid region of human immunodeficiency virus type 1 that affect particle formation and gag-pol precursor incorporation. *J. Virol.* **69**, 6106-6114.
- Starcich, B. R., Hahn, B. H., Shaw, G. M., McNeely, P. D., Modrow, S., Wolf, H., Parks, E. S., Parks, W. P., Josephs, S. F., Gallo, R. C., and Wong Staal, F. (1986). Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell* **45**, 637-648.
- Stellbrink, H. J., Zoller, B., Fenner, T., Lichtenberg, G., Vanlunzen, J., Albrecht, H., Thiele, H. G., and Greten, H. (1996). Rapid plasma virus and CD4+ T-cell turnover in HIV-1 infection: Evidence for an only transient interruption by treatment. *AIDS* **10**, 849-857.
- Stemmer, W. P. (1994a). DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. U.S.A* **91**, 10747-10751.
- Stemmer, W. P. (1994b). Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389-391.
- Steuler, H., Storch Hagenlocher, B., and Wildemann, B. (1992). Distinct populations of human immunodeficiency virus type 1 in blood and cerebrospinal fluid. *AIDS. Res. Hum. Retroviruses.* **8**, 53-59.
- Stewart, S. A., Poon, B., Jowett, J. B. M., and Chen, I. S. Y. (1997). Human immunodeficiency virus type 1 vpr induces apoptosis following cell cycle arrest. *J. Virol.* **71**, 5579-5592.
- Strebel, K., Klimkait, T., Maldarelli, F., and Martin, M. A. (1989). Molecular and biochemical analysis of HIV-1 vpu protein. *J. Virol.* **63**, 3784-3791.
- Strizki, J. M., Albright, A. V., Sheng, H., O'Connor, M., Perrin, L., and Gonzalezscarano, F. (1996). Infection of primary human microglia and monocyte-derived macrophages with human immunodeficiency virus type 1 isolates: Evidence of differential tropism. *J. Virol.* **70**, 7654-7662.
- Stromberg, K., Hurley, N. E., Davis, N. L., Rueckert, R. R., and Fleissner, E. (1974). Structural studies of avian myeloblastosis virus: comparison of polypeptides in virion and core component by dodecyl sulfate- polyacrylamide gel electrophoresis. *J. Virol.* **13**, 513-528.
- Stuhlmann, H., Dieckmann, M., and Berg, P. (1990). Transduction of cellular neo mRNA by retrovirus-mediated recombination. *J. Virol.* **64**, 5783-5796.
- Swingler, S., Mann, A., Jacque, J. M., Brichacek, B., Sasseville, V. G., Williams, K., Lackner, A. A., Janoff, E. N., Wang, R., Fisher, D., and Stevenson, M. (1999). HIV-1 Nef mediates lymphocyte chemotaxis and activation by infected macrophages. *Nature Med* **5**, 997-1003.

- Sykes, K. F. and Johnston, S. A. (1999). Linear expression elements: a rapid, in vivo, method to screen for gene functions. *Nat. Biotechnol.* **17**, 355-359.
- Taddeo, B., Federico, M., Titti, F., Rossi, G. B., and Verani, P. (1993). Homologous superinfection of both producer and nonproducer HIV- infected cells is blocked at a late retrotranscription step. *Virology* **194**, 441-452.
- Takehisa, J., Zekeng, L., Ido, E., YamaguchiKabata, Y., Mboudjeka, I., Harada, Y., Miura, T., Kaptue, L., and Hayami, M. (1999). Human immunodeficiency virus type 1 intergroup (M/O) recombination in Cameroon. *J Virol* **73**, 6810-6820.
- Takeuchi, Y., Akutsu, M., Murayama, K., Shimizu, N., and Hoshino, H. (1991). 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**, 1710-1718.
- Tamm, L. K. and Han, X. (2000). Viral fusion peptides: a tool set to disrupt and connect biological membranes. *Biosci. Rep.* **20**, 501-518.
- Tan, K. M., Liu, J. H., Wang, J. H., Shen, S., and Lu, M. (1997). Atomic structure of a thermostable subdomain of HIV-1 gp41. *Proc Natl Acad Sci USA* **94**, 12303-12308.
- Tang, Y. W., Huong, J. T., Lloyd, R. M., Jr., Spearman, P., and Haas, D. W. (2000). Comparison of human immunodeficiency virus type 1 RNA sequence heterogeneity in cerebrospinal fluid and plasma. *Journal of Clinical Microbiology* **38**, 4637-4639.
- Temin, H. M. (1989). Is HIV unique or merely different? *J. Acquir. Immune. Defic. Syndr.* **2**, 1-9.
- Temin, H. M. (1993). Retrovirus variation and reverse transcription - abnormal strand transfers result in retrovirus genetic variation. *Proc. Natl. Acad. Sci. USA* **90**, 6900-6903.
- Temin, H. M. and Mizutani, S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Biotechnology* **24**, 51-56.
- Tersmette, M., de Goede, R. E. Y., Al, B. J. M., Winkel, I. N., Gruters, R. A., Cuypers, H. T. M., Huisman, H. G., and Miedema, F. (1988). Differential syncytium-inducing capacity of human immunodeficiency virus isolates: frequent detection of syncytium- inducing isolates in patients with acquired immunodeficiency syndrome (AIDS) and AIDS-related complex. *J. Virol.* **62**, 2026-2032.
- Thomson, M. M., Perez-Alvarez, L., and Najera, R. (2002). Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. *Lancet Infect. Dis.* **2**, 461-471.
- Tremblay, M. J., Fortin, J. F., and Cantin, R. (1998). The acquisition of host-encoded proteins by nascent HIV-1. *Immunol. Today* **19**, 346-351.
- Tristem, M., Purvis, A., and Quicke, D. L. J. (1998). Complex evolutionary history of primate lentiviral vpr genes. *Virology* **240**, 232-237.
- Trkola, A., Dragic, T., Arthos, J., Binley, J. M., Olson, W. C., Allaway, G. P., Chengmayer, C., Robinson, J., Maddon, P. J., and Moore, J. P. (1996). CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. *Nature* **384**, 184-187.
- Tscherning, C., Alaeus, A., Fredriksson, R., Bjorndal, A., Deng, H. K., Littman, D. R., Fenyo, E. M., and Albert, J. (1998). Differences in chemokine coreceptor usage between genetic subtypes of HIV-1. *Virology* **241**, 181-188.
- Ugolini, S., Mondor, I., and Sattentau, Q. J. (1999). HIV-1 attachment: another look. *Trends Microbiol.* **7**, 144-149.

- Valentin, A., Lu, W. H., Rosati, M., Schneider, R., Albert, J., Karlsson, A., and Pavlakis, G. N. (1998). Dual effect of interleukin 4 on HIV-1 expression: Implications for viral phenotypic switch and disease progression. *Proc Natl Acad Sci USA* **95**, 8886-8891.
- Valentin, A., Rosati, M., Patenaude, D. J., Hatzakis, A., Kostrikis, L. G., Lazanas, M., Wyvill, K. M., Yarchoan, R., and Pavlakis, G. N. (2002). Persistent HIV-1 infection of natural killer cells in patients receiving highly active antiretroviral therapy. *Proc.Natl.Acad.Sci.U.S.A* **99**, 7015-7020.
- Vallee, H. and Carre, H. (1904). Sur la nature infectieuse de l'anemi du cheval. *C.R.Hebd.Seances Acad.Sci.Ser.D Sci.Nat* **139**, 331-333.
- Vantwout, A. B., Ran, L. J., Kuiken, C. L., Kootstra, N. A., Pals, S. T., and Schuitemaker, H. (1998). Analysis of the temporal relationship between human immunodeficiency virus type 1 quasispecies in sequential blood samples and various organs obtained at autopsy. *J Virol* **72**, 488-496.
- Vanwamel, J. L. B. and Berkhout, B. (1998). The first strand transfer during HIV-1 reverse transcription can occur either intramolecularly or intermolecularly (Rapid Communication). *Virology* **244**, 245-251.
- Vartanian, J. P., Meyerhans, A., Asjo, B., and Wain Hobson, S. (1991). Selection, recombination, and G-A hypermutation of human immunodeficiency virus type 1 genomes. *J.Virol* **65**, 1779-1788.
- Vazeux, R. (1991). AIDS encephalopathy and tropism of HIV for brain monocytes/macrophages and microglial cells. *Pathobiology*. **59**, 214-218.
- Vella, S. and Palmisano, L. (2000). Antiretroviral therapy: state of the HAART. *Antiviral Res.* **45**, 1-7.
- Verani, A., Pesenti, E., Polo, S., Tresoldi, E., Scarlatti, G., Lusso, P., Siccardi, A. G., and Vercelli, D. (1998). CXCR4 is a functional coreceptor for infection of human macrophages by CXCR4-dependent primary HIV-1 isolates. *J Immunol* **161**, 2084-2088.
- Veronese, F. D., DeVico, A. L., Copeland, T. D., Oroszlan, S., Gallo, R. C., and Sarngadharan, M. G. (1985). Characterization of gp41 as the transmembrane protein coded by the HTLV-III/LAV envelope gene. *Science*. **229**, 1402-1405.
- Viguera, E., Canceill, D., and Ehrlich, S. D. (2001). In vitro replication slippage by DNA polymerases from thermophilic organisms. *J.Mol.Biol.* **312**, 323-333.
- Vincent, M. J., Raja, N. U., and Jabbar, M. A. (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.
- Vogt, P. K. (1971). Genetically stable reassortment of markers during mixed infection with avian tumor viruses. *Virology* **46**, 947-952.
- Vogt, V. M. (1997). Retroviral Virions and Genomes. In "Retroviruses" (J. M. Coffin, S. H. Hughes, and H. E. Varmus, Eds.), pp. 27-70. Cold Spring Harbor Laboratory Press, New York.
- Vonschwedler, U., Kornbluth, R. S., and Trono, D. (1994). The nuclear localization signal of the matrix protein of human immunodeficiency virus type 1 allows the establishment of infection in macrophages and quiescent t lymphocytes. *Proc.Natl.Acad.Sci.USA* **91** , 6992-6996.
- Vonschwedler, U., Song, J. P., Aiken, C., and Trono, D. (1993). vif is crucial for human immunodeficiency virus type-1 proviral DNA synthesis in infected cells. *J.Virol.* **67**, 4945-4955.

- Voulgaropoulou, F., Tan, B., Soares, M., Hahn, B., and Ratner, L. (1999). Distinct human immunodeficiency virus strains in the bone marrow are associated with the development of thrombocytopenia. *J. Virol.* **73**, 3497-3504.
- Waga, S. and Stillman, B. (1998). The DNA replication fork in eukaryotic cells. *Annu. Rev. Biochem.* **67**, 721-751.
- Wainhobson, 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.
- Wang, B., Lal, R. B., Dwyer, D. E., Miranda-Saksena, M., Boadle, R., Cunningham, A. L., and Saksena, N. K. (2000). Molecular and biological interactions between two HIV-1 strains from a coinfecting patient reveal the first evidence in favor of viral synergism. *Virology* **274**, 105-119.
- Wang, G. C. and Wang, Y. (1997). Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. *Appl. Environ. Microbiol.* **63**, 4645-4650.
- Wang, R. F. and Mullins, J. I. (1995). Mammalian cell/vaccinia virus expression vectors with increased stability of retroviral sequences in *Escherichia coli*: production of feline immunodeficiency virus envelope protein. *Gene* **153**, 197-202.
- Wang, T. H., Donaldson, Y. K., Brettell, R. P., Bell, J. E., and Simmonds, P. (2001). Identification of shared populations of human immunodeficiency virus type 1 infecting microglia and tissue macrophages outside the central nervous system. *J. Virol.* **75**, 11686-99.
- Wang, Z. X., Berson, J. F., Zhang, T. Y., Cen, Y. H., Sun, Y., Sharron, M., Lu, Z. H., and Peiper, S. C. (1998). CXCR4 sequences involved in coreceptor determination of human immunodeficiency virus type-1 tropism - Unmasking of activity with M- tropic Env glycoproteins. *J Biol Chem* **273**, 15007-15015.
- Watkins, B. A., Dorn, H. H., Kelly, W. B., Armstrong, R. C., Potts, B. J., Michaels, F., Kufta, C. V., and Dubois Dalcq, M. (1990). Specific tropism of HIV-1 for microglial cells in primary human brain cultures. *Science.* **249**, 549-553.
- Wei, Q. and Fultz, P. N. (1998). Extensive diversification of human immunodeficiency virus type 1 subtype B strains during dual infection of a chimpanzee that progressed to AIDS. *J Virol* **72**, 3005-3017.
- Wei, X. P., Ghosh, S. K., Taylor, M. E., Johnson, V. A., Emini, E. A., Deutsch, P., Lifson, J. D., Bonhoeffer, S., Nowak, M. A., Hahn, B. H., Saag, M. S., and Shaw, G. M. (1995). Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **373**, 117-122.
- Weiss, C. D., Levy, J. A., and White, J. M. (1990). Oligomeric organization of gp120 on infectious human immunodeficiency virus type 1 particles. *J. Virol.* **64**, 5674-5677.
- Weissman, D., Rabin, R. L., Arthos, J., Rubbert, A., Dybul, M., Swofford, R., Venkatesan, S., Farber, J. M., and Fauci, A. S. (1997). Macrophage-tropic HIV and SIV envelope proteins induce a signal through the CCR5 chemokine receptor. *Nature* **389**, 981-985.
- Wieggers, K., Rutter, G., Schubert, U., Grattinger, M., and Krausslich, H. G. (1999). Cyclophilin A incorporation is not required for human immunodeficiency virus type 1 particle maturation and does not destabilize the mature capsid. *Virology* **257**, 261-274.

- Wild, C. T., Shugars, D. C., Greenwell, T. K., Mcdanal, C. B., and Matthews, T. J. (1994). Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection. *Proc.Natl.Acad.Sci.USA* **91**, 9770-9774.
- Wiley, C. A., Schrier, R. D., Nelson, J. A., Lampart, P. W., and Oldstone, M. B. A. (1986). Cellular location of human immunodeficiency virus infection within the brains of acquired immune deficiency syndrome patients. *Proc.Natl.Acad.Sci.USA* **83**, 7089-7093.
- Wilkinson, D. A., Operskalski, E. A., Busch, B. P., Mosley, J. W., and Koup, R. A. (1998). A 32-bp deletion within the CCR5 locus protects against transmission of parenterally acquired human immunodeficiency virus but does not affect progression to AIDS-defining illness. *J Infect Dis* **178**, 1163-1166.
- Wiley, R. L., Bonifacino, J. S., Potts, B. J., Martin, M. A., and Klausner, R. D. (1988). Biosynthesis, cleavage, and degradation of the human immunodeficiency virus 1 envelope glycoprotein gp160. *Proc.Natl.Acad.Sci.U.S.A* **85**, 9580-9584.
- Winters, M. A., Schapiro, J. M., Lawrence, J., and Merigan, T. C. (1998). Human immunodeficiency virus type 1 protease genotypes and in vitro protease inhibitor susceptibilities of isolates from individuals who were switched to other protease inhibitors after long-term saquinavir treatment. *J Virol* **72**, 5303-5306.
- Wolfs, T. F. W., de Jong, J. J., Van den Berg, H., Tijnagel, J. M., Krone, W. J., and Goudsmit, J. (1990). Evolution of sequences encoding the principal neutralization epitope of human immunodeficiency virus 1 is host dependent, rapid, and continuous. *Proc.Natl.Acad.Sci.U.S.A* **87**, 9938-9942.
- Wolinsky, S. M., Korber, B. T. M., Neumann, A. U., Daniels, M., Kunstman, K. J., Whetsell, A. J., Furtado, M. R., Cao, Y. Z., Ho, D. D., Safrit, J. T., and Koup, R. A. (1996). Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science* **272**, 537-542.
- Wolinsky, S. M., Wike, C. M., Korber, B. T. M., Hutto, C., Parks, W. P., Rosenblum, L. A., Kunstman, K. J., Furtado, M. R., and Munoz, J. L. (1992). Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. *Science* **255**, 1134-1137.
- Wong, J., Ignacio, C., Torriani, F., Havlir, D., Fitch, N. J. S., and Richman, D. D. (1997). In vivo compartmentalization of human immunodeficiency virus: Evidence from the examination of *pol* sequences from autopsy tissues. *J.Virol.* **71**, 2059-2071.
- Wong, P. K. and McCarter, J. A. (1973). Genetic studies of temperature-sensitive mutants of Moloney-murine leukemia virus. *Virology* **53**, 319-326.
- Wu, L. J., Gerard, N. P., Wyatt, R., Choe, H., Parolin, C., Ruffing, N., Borsetti, A., Cardoso, A. A., Desjardin, E., Newman, W., Gerard, C., and Sodroski, J. (1996). CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature* **384**, 179-183.
- Wu, W. M., Blumberg, B. M., Fay, P. J., and Bambara, R. A. (1995). Strand transfer mediated by human immunodeficiency virus reverse transcriptase in vitro is promoted by pausing and results in misincorporation. *Journal of Biological Chemistry* **270**, 325-332.
- Wyatt, R., Kwong, P. D., Desjardins, E., Sweet, R. W., Robinson, J., Hendrickson, W. A., and Sodroski, J. G. (1998). The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **393**, 705-711.
- Wyatt, R., Sullivan, N., Thali, M., Repke, H., Ho, D., Robinson, J., Posner, M., and Sodroski, J. (1993). Functional and immunologic characterization of human immunodeficiency virus type-1 envelope glycoproteins containing deletions of the major variable regions. *J.Virol.* **67**, 4557-4565.

- Yachou, A. and Sekaly, R. P. (1999). Binding of soluble recombinant HIV envelope glycoprotein, rgp120, induces conformational changes in the cellular membrane-anchored CD4 molecule. *Biochemical and Biophysical Research Communications* **265**, 428-433.
- Yamada, K., Morozumi, H., and Okamoto, T. (1995). LTR-directed homologous recombination of full-length HIV-1 provirus clone in *reca* (-) bacteria. *Arch. Virol.* **140**, 1007-1014.
- Yamaguchi-Kabata, Y. and Gojobori, T. (2000). Reevaluation of amino acid variability of the human immunodeficiency virus type 1 gp120 envelope glycoprotein and prediction of new discontinuous epitopes. *J. Virol.* **74**, 4335-4350.
- Yang, Y. L., Wang, G. Q., Dorman, K., and Kaplan, A. H. (1996). Long polymerase chain reaction amplification of heterogeneous HIV type 1 templates produces recombination at a relatively high frequency. *AIDS Res. Hum. Retroviruses* **12**, 303-306.
- Yarchoan, R., Perno, C. F., Thomas, R. V., Collins, J. M., Myers, C. E., and Broder, S. (1988). Phase 1 studies of 2',3'-dideoxycytidine in severe HIV as a single agent and altering with zidovudine (AZT). *Lancet* **i**, 76-81.
- Yeh, M. W., Kaul, M., Zheng, J., Nottet, H. S., Thylin, M., Gendelman, H. E., and Lipton, S. A. (2000). Cytokine-stimulated, but not HIV-infected, human monocyte-derived macrophages produce neurotoxic levels of l-cysteine. *J. Immunol.* **164**, 4265-4270.
- Yi, Y. J., Rana, S., Turner, J. D., Gaddis, N., and Collman, R. G. (1998). CXCR-4 is expressed by primary macrophages and supports CCR5- independent infection by dual-tropic but not T-tropic isolates of human immunodeficiency virus type 1. *J Virol* **72**, 772-777.
- Yu, H., Jetzt, A. E., Ron, Y., Preston, B. D., and Dougherty, J. P. (1998). The nature of human immunodeficiency virus type 1 strand transfers. *J Biol Chem* **273**, 28384-28391.
- Zekeng, L., Gurtler, L., Ze, E. A., Samabbenyi, A., Mbouniessomba, G., Mpoudingolle, E., Monnylobe, M., Tapko, J. B., and Kaptue, L. (1994). Prevalence of HIV-1 subtype o infection in cameroon: preliminary results. *AIDS* **8**, 1626-1628.
- Zennou, V., Petit, C., Guetard, D., Nerhbass, U., Montagnier, L., and Charneau, P. (2000). HIV-1 genome nuclear import is mediated by a central DNA flap. *Cell* **101**, 173-185.
- Zennou, V., Serguera, C., Sarkis, C., Colin, P., Perret, E., Mallet, J., and Charneau, P. (2001). The HIV-1 DNA flap stimulates HIV vector-mediated cell transduction in the brain. *Nat. Biotechnol.* **19**, 446-450.
- Zhang, G. H., Zapp, M. L., Yan, G. C., and Green, M. R. (1996). Localization of HIV-1 RNA in mammalian nuclei. *J. Cell Biol.* **135**, 9-18.
- Zhang, H., Yang, B., Pomerantz, R. J., Zhang, C., Arunachalam, S. C., and Gao, L. (2003). The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* **424**, 94-98.
- Zhang, J. and Temin, H. M. (1993). Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication. *Science* **259**, 234-238.
- Zhang, J. and Temin, H. M. (1994). Retrovirus recombination depends on the length of sequence identity and is not error prone. *J. Virol.* **68**, 2409-2414.
- Zhang, L., Rowe, L., He, T., Chung, C., Yu, J., Yu, W., Talal, A., Markowitz, M., and Ho, D. D. (2002). Compartmentalization of surface envelope glycoprotein of human immunodeficiency virus type 1 during acute and chronic infection. *J Virol* **76**, 9465-9473.

- Zhang, L. Q., MacKenzie, P., Cleland, A., Holmes, E. C., Leigh Brown, A. J., and Simmonds, P. (1993). Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. *J. Virol.* **67**, 3345-3356.
- Zhao, L. J., Wang, L. L., Mukherjee, S., and Narayan, O. (1994). Biochemical mechanism of HIV-1 vpr function - oligomerization mediated by the n-terminal domain. *Journal of Biological Chemistry* **269**, 32131-32137.
- Zheng, R. L., Jenkins, T. M., and Craigie, R. (1996). Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc Natl Acad Sci USA* **93**, 13659-13664.
- Zhu, T., Mo, H., Wang, N., Nam, D. S., Cao, Y., Koup, R. A., and Ho, D. (1993a). Genotypic and phenotypic characterization of HIV-1 patients with primary infection. *Science*. **261**, 1179-1181.
- Zhu, T. F., Korber, B. T., Nahmias, A. J., Hooper, E., Sharp, P. M., and Ho, D. D. (1998). An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* **391**, 594-597.
- Zhu, T. F., Mo, H. M., Wang, N., Nam, D. S., Cao, Y. Z., Koup, R. A., and Ho, D. D. (1993b). Genotypic and phenotypic characterization of HIV-1 in patients with primary infection. *Science* **261**, 1179-1181.
- Zhu, T. F., Wang, N., Carr, A., Nam, D. S., MoorJankowski, R., Cooper, D. A., and Ho, D. D. (1996). 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**, 3098-3107.
- Zhuang, J., Jetzt, A. E., Sun, G., Yu, H., Klarmann, G., Ron, Y., Preston, B. D., and Dougherty, J. P. (2002). Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J Virol* **76**, 11273-11282.
- Zybarth, G. and Carter, C. (1995). Domains upstream of the protease (PR) in human immunodeficiency virus type 1 gag-pol influence PR autoprocessing. *J. Virol.* **69**, 3878-3884.
- Zylstra, P., Rothenfluh, H. S., Weiller, G. F., Blanden, R. V., and Steele, E. J. (1998). PCR amplification of murine immunoglobulin germline V genes: strategies for minimization of recombination artefacts. *Immunol Cell Biol.* **76**, 395-405.

Appendix

Mosaic Structure of the Human Immunodeficiency Virus Type 1 Genome Infecting Lymphoid Cells and the Brain: Evidence for Frequent In Vivo Recombination Events in the Evolution of Regional Populations

A. MORRIS,¹ M. MARSDEN,¹ K. HALCROW,¹ E. S. HUGHES,¹ R. P. BRETTELE,²
J. E. BELL,³ AND P. SIMMONDS^{1*}

Department of Medical Microbiology, University of Edinburgh, Edinburgh EH8 9AG,¹ and Regional Infectious Diseases Unit, Western General Hospital,² and Department of Neuropathology, University of Edinburgh, Western General Hospital,³ Edinburgh EH4 2XU, United Kingdom

Received 16 December 1998/Accepted 7 July 1999

In addition to immunodeficiency, human immunodeficiency virus type 1 (HIV-1) can cause cognitive impairment and dementia through direct infection of the brain. To investigate the adaptive process and timing of HIV-1 entry into the central nervous system, we carried out an extensive genetic characterization of variants amplified from different regions of the brain and determined their relatedness to those in lymphoid tissue. HIV-1 genomes infecting different regions of the brain of one study subject with HIV encephalitis (HIVE) had a mosaic structure, being assembled from different combinations of evolutionarily distinct lineages in p17^{gag}, pol, individual hypervariable regions of gp120 (V1/V2, V3, V4, and V5), and gp41/nef. Similar discordant phylogenetic relationships were observed between p17^{gag} and V3 sequences of brain and lymphoid tissue from three other individuals with HIVE. The observation that different parts of the genome of HIV infecting a particular tissue can have different evolutionary histories necessarily limits the conclusions that can be drawn from previous studies of the compartmentalization of distinct HIV populations in different tissues, as these have been generally restricted to sequence comparisons of single subgenomic regions. The complexity of viral populations in the brain produced by recombination could provide a powerful adaptive mechanism for the spread of virus with new phenotypes, such as antiviral resistance or escape from cytotoxic T-cell recognition into existing tissue-adapted virus populations.

Isolates of human immunodeficiency virus (HIV) show great heterogeneity in their kinetics of replication, coreceptor usage, cellular tropism, and cytopathic effects. Differences in these properties have been hypothesized to underlie in vivo differences in pathogenicity, which in turn may influence the rate of disease progression in HIV-infected individuals (1, 2, 5, 9, 11, 16, 19, 20, 28, 40, 44). Differences in phenotype may also underlie differences in in vivo cellular tropism, which would substantiate the hypothesis that the different populations of HIV infecting lymphoid tissue, brain, and other tissues may have originated through an adaptive process following primary infection. Difficulties in recovering HIV from nonlymphoid tissues have to date prevented extensive analysis of their biological properties, although the existence of consistent sequence differences in the envelope gene from those recovered from lymphoid tissues provides indirect evidence for specific cellular tropisms (3, 13, 14, 17, 22, 26, 27, 30, 31, 34–36, 42). However, an alternative hypothesis proposes that differences in the rate of virus turnover in different cell types may lead to the observed population differences, given the rapid temporal change in HIV populations over time in peripheral blood mononuclear cells, in lymph nodes (LN), and among HIV variants recovered from the gastrointestinal tract (27, 45, 46).

To investigate whether differences in V3 and elsewhere in the HIV genome reflect tissue adaptation, or whether they

arise simply through limited spatial or temporal sampling, we have compared nucleotide sequences in different regions of the HIV type 1 (HIV-1) genome from lymphoid tissue with autopsy samples from anatomically separated parts of the brains of four study subjects with HIV encephalitis (HIVE). We also determined whether sequences in p17^{gag}, pol, and different regions of the env gene (V1/V2, V3, V4, V5, and gp41/nef) between different brain samples and those from lymphoid tissue provided equivalent evidence for tissue-specific compartmentalization of HIV-1.

MATERIALS AND METHODS

Study subjects. Frozen samples of LN or spleen and from several anatomically distinct regions of the brain from four individuals with HIV giant cell encephalitis were stored at autopsy (risk group, CD4 counts, and brain pathology are summarized in Table 1). Subject NA129 had received zidovudine monotherapy for 17 months up to approximately 1 year before death; ddC was used for 1 month, finishing 3 months before death. The other study subjects had received minimal antiviral treatment: for NA234, a single course of zidovudine for a duration of 1 month at 1 year before death; for NA021, zidovudine intermittently over 1 year at 5 years before death and zidovudine-ddC for 1 month at 1 year before death; for NA173, zidovudine for 4 months at 2 years before death and then for 1 month at 1 year before death. None of the study subjects showed evidence for genotypic resistance to zidovudine or other antiviral agents (42a).

Pathology examination. The brains were examined pathologically as previously described (4). Assessment of pathology findings was undertaken blind to the PCR analysis and validated by three independent observers.

DNA extraction and amplification. DNA was extracted from the brain, LN, and spleen and quantified as previously described (15). Total DNA was quantified spectrophotometrically, while HIV proviral DNA was semiquantitated by amplification, using the p17^{gag} primers, of serial 10-fold dilutions of DNA, with the last positive dilution used to indicate the minimum proviral load in the sample. Samples were used for sequence comparison only if proviral frequencies were >100 copies/10⁶ cells, and this excluded analysis of left parietal (LP) and both cerebellum samples from NA234. Low levels were detected in two atrophied LN

* Corresponding author. Mailing address: Department of Medical Microbiology, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, United Kingdom. Phone: 44 131 650 3138. Fax: 44 131 650 6531. E-mail: Peter.Simmonds@ed.ac.uk.

TABLE 1. Clinical background and pathology appearance of study subjects

Study subject	Age (yr)	Risk group ^a	CD4/ μ l		HIV-related pathology ^b						Other pathology
			1 yr before death	3 mo before death	PV	WM	GM	BG	BS	Cere	
NA173	34	MH	1	0	++	+/-	+/-	+/-	+/-	-	Toxoplasma
NA021	49	IVDU	300	87	+	++	+	+	+/-	+/-	
NA128	48	MH	12	4	+/-	+++	++	++	++	-	Toxoplasma
NA234	28	IVDU	90	30	+	++	++	++	+	+	Lymphoma

^a MH, male homosexual; IVDU, intravenous drug user.

^b Scored as - (negative), +/- (sparse pathology), +, (slight pathology), or +++ (severe pathology in different brain compartments). PV, perivascular; WM, white matter; GM, grey matter; BG, basal ganglia; BS, brain stem; Cere, cerebellum.

samples from NA128, and lymphoid sequences were therefore obtained from the spleen. For nucleotide sequencing, 1- μ g aliquots of extracted DNA were amplified as previously described (39) or amplified after dilution to an endpoint (see below).

Amplified DNA was either sequenced directly by cycle sequencing (Amersham) or cloned into pGEM, using poly(T) overhangs (pGEM-T Vector system; Promega). Miniprep DNA from clones was sequenced by using a Sequenase version 2.0 kit (U.S. Biochemical) according to the manufacturer's protocol.

Nucleotide sequencing. For samples with evidence of sequence heterogeneity, amplified DNA was cloned and approximately 10 clones were sequenced. Sequences from other samples without evidence for heterogeneity (no detectable multiple bands at any position on a sequencing gel lane) or length polymorphism analysis (23) were directly sequenced. The sequence data set for the p17^{gag} region of NA234 was assembled from cloned sequences from each of the brain regions and extended from positions 405 to 795 in the HIV_{LAI} genomic sequence (GenBank accession no. K02013). Sequences in the *pol* region of NA234 were directly sequenced from amplified DNA and compared between bases 2189 and 2842, using the PCR primers used for detection of antiviral resistance mutations (43); there was no evidence for intrasample sequence heterogeneity. Sequences in the V1/V2 region from NA234 were amplified by using primers as previously described (23). Amplified DNA was homogeneous by length polymorphism analysis, and nucleotide sequences from regions other than LN, left frontal (LF), and right frontal (RF) were obtained by direct sequencing of amplified DNA. Sequences in V3 and V4/V5 were amplified by using primers as previously described (38). Length polymorphisms in the V4/V5 region in some brain samples from NA234 necessitated multiple clones to be sequenced. Sequences from each tissue were amplified in the gp41/*nef* region by using outer primers 5'-AGGGC TGCTATTAACAAGAGATGG-3' and 5'-GTAGCTGCTGTATTGCTACTT GTG-3' and inner primers 5'-AGGCTTCAGACCTGGAGGAGG-3' and 5'-TATGCTACTTGTGATTGCTCCATG-3' (5' base positions 7166, 8541, 7215, and 8531, respectively). Direct sequencing was carried out on the amplified DNA from the 3' end by using the antisense inner primer, while an internal primer (5'-TGAGGGGACAGATAGGGTTATAGA-3', position of 5' base 8287 in HIV_{LAI} genome) was used to sequence the 5' end.

Sequences were aligned and distances were estimated with the Simmonic 2000 Sequence Editor package. Synonymous and nonsynonymous distances and standard errors were estimated by the method of Nei and Gojobori (33). Phylogenetic analysis was carried out with the MEGA program (29). The nucleotide sequences from p17^{gag} and V3 amplified from each of the study subjects were compared with each other and with a range of standard HIV-1 variants. Each set of sequences from the four study subjects was monophyletic in both genomic regions and distinct from those of the published sequences of subtype B: HIVSF2 (K02007), HIVRF (M17451), HIVOYI (M26727), HIVLAI (K02013), HIVJRF1 (M74978), HIVYU2 (M93258), HIVCAM1 (D10112), HIVNY5CG (M38431), HIVHAN (U43131), HIVWMJ22 (M12507), and HIVSFAAA (M65024). This comparison provided no evidence for coinfection with more than one epidemiologically unrelated HIV strains or for intersample or exogenous laboratory contamination.

Nucleotide sequence accession numbers. Nucleotide sequences obtained in this study have been submitted to GenBank and assigned accession no. AF174692 through AF175123.

RESULTS

Sequence relationships in different regions of the HIV-1 genome. Autopsy samples were obtained from multiple regions of the brain and from LN of an individual (NA234) with autopsy evidence of HIV. The p17^{gag} region was amplified by nested PCR and cloned from samples with virus loads of greater than 100 proviral copies/10⁶ cells. These nucleotide sequences formed two distinct evolutionary lineages, each with bootstrap support of $\geq 80\%$. Sequences from LN, brain stem,

right parietal (RP) and LF regions (lineage A) showed a closer sequence relationship to each other than to variants obtained from elsewhere in the brain (lineage B) (Fig. 1). At synonymous sites, the mean pairwise Jukes-Cantor (J-C) distances were 0.057 (mean standard error ± 0.026) among members of lineage A and 0.055 (± 0.024) for lineage B, almost nonoverlapping with the range of pairwise distances observed between the two lineages (mean synonymous distance 0.109 ± 0.038). On the basis of the previously established mean rate of sequence change in this region of *gag* (0.6 to 0.7% per site per year [24, 25]), these distances suggest a time of divergence between lineages A and B of around 8.4 ± 2.9 years, compared with a likely duration of HIV infection of around 10 years in this study subject.

In contrast to LN sequences, those from other regions of the brain were relatively homogeneous, apart from those from the RP region, which formed at least three separate clusters within lineage A, and a single sequence from the LF region, which grouped in lineage B instead of lineage A. Sequences in lineage B, comprising exclusively brain-derived variants, clustered by tissue origin, with a tendency for variants recovered from adjacent tissues to be more similar to each other (e.g., left occipital [LO] and right occipital [RO] regions) than to variants from virus with greater physical separation (e.g., occipital region to RF region). Most regional variants in lineage B had common ancestors distinct from any variants found in lymphoid tissue, and more recent times of divergence than that between lineages A and B can be inferred. Variants within lineage B may therefore have originated from the spread of HIV within the brain rather than from multiple seeding from the peripheral circulation and thus differ in origin from brain-derived variants in lineage A (LF and RP). To confirm the separate groupings of LF and RF sequences, DNAs extracted from these tissues and from LN were separately analyzed by limiting dilution as previously described (39). Sequences obtained were similar or identical to those obtained by cloning of PCR products (data not shown).

In marked contrast to the relationships observed in p17^{gag}, LF and RF sequences from both V1/V2 and V3 regions in gp120 grouped together in a distinct lineage from those of LN sequences (bootstrap support $\geq 80\%$) (Fig. 2A and B), demonstrating that variants from LF and RF regions shared a common ancestor distinct from that of variants in the LN region. The probability of this discordant phylogeny arising through sampling was < 0.0001 (Fisher's exact test). To investigate the robustness of the difference in branching order in p17^{gag} and V3, a user-defined tree of p17^{gag} sequences with the branching order of V3 was created by using RETREE in the PHYLIP package (18). With the Hasegawa-Kishino-Yano test, a p17^{gag} tree with the V3 branching order was significantly less

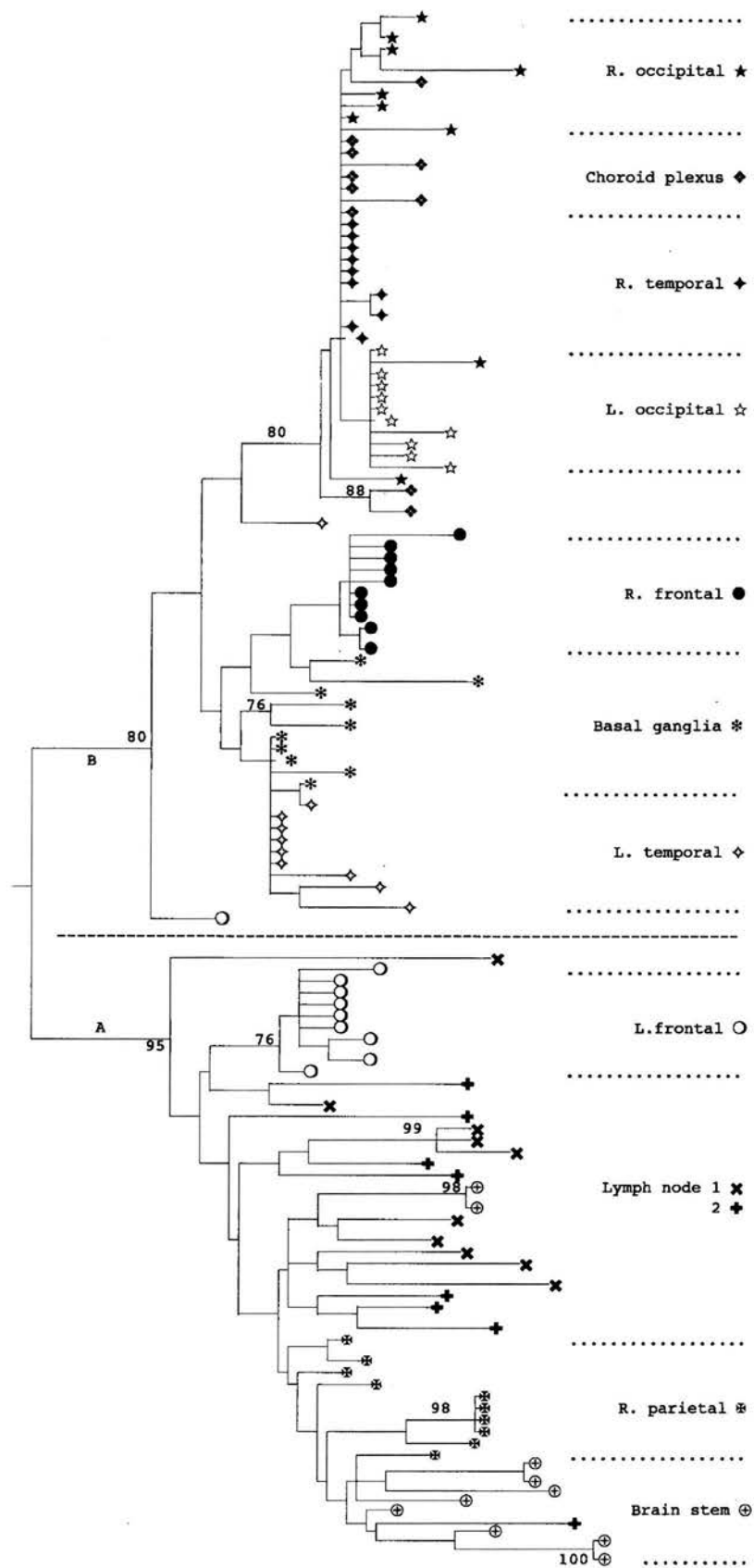


FIG. 1. Phylogenetic analysis of the p17^{gag} region amplified from different regions of the brain and from lymphoid tissue of NA234. Divergence between nucleotide sequences was estimated from J-C distances (scale indicated below tree), and the tree was constructed from the distance matrix by the neighbor-joining method. The robustness of groupings was indicated by bootstrap resampling of 100 data sets, with values of $\geq 75\%$ indicated on branches; lineages are indicated by single letters. The tree was rooted by using the sequence of HIV_{LAI} (accession no. K02013) as an outgroup.

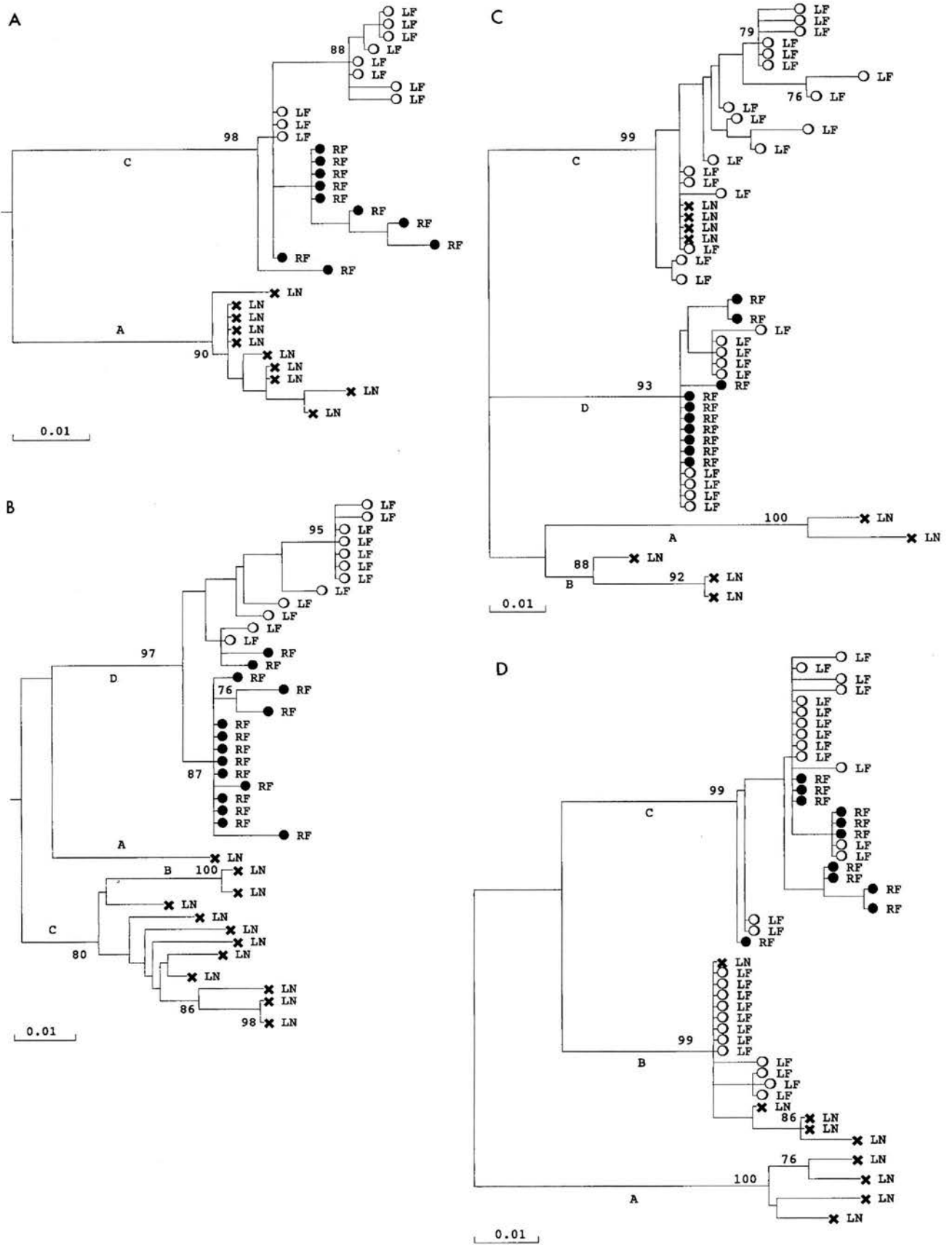


FIG. 2. Comparison of sequences from LF and RF regions of the brain of NA234 with those of the LN region in V1/V2 (A), V3 (B), V4 (C), and V5 (D) hypervariable regions (outgroup, symbols, and bootstrap method as for Fig. 1).

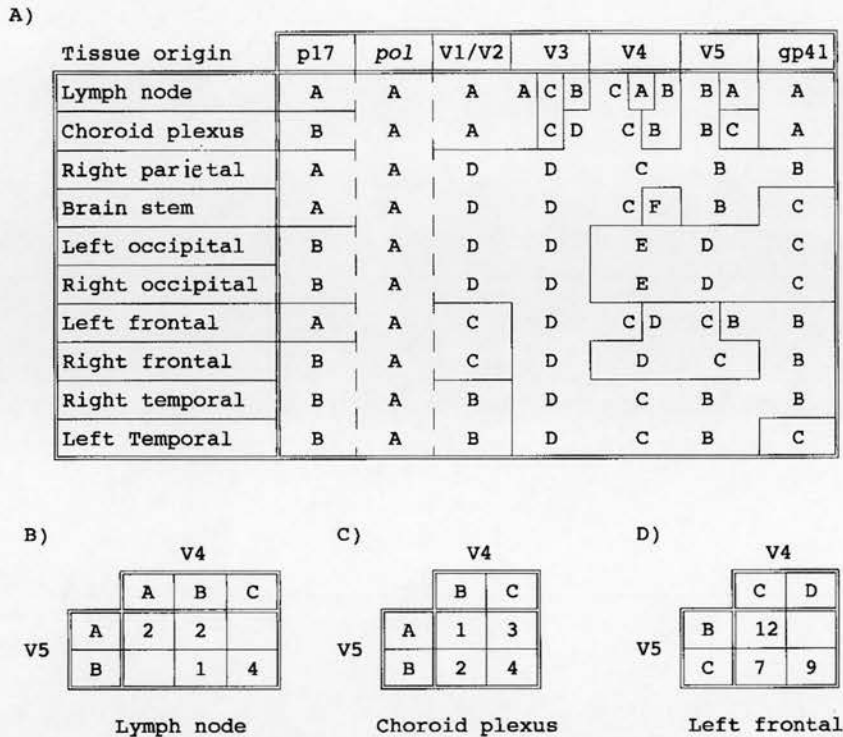


FIG. 3. (A) Distribution of phylogenetic groupings in different regions of the genome of HIV-1 variants infecting NA234. Letters refer to bootstrap-supported groupings identified by phylogenetic analysis (Fig. 1, 2, and 3); vertical bars indicate discontinuities in groupings that require the minimum number of recombination events. The sequence conservation and monophylogeny of the *pol* region originated either through recombination or from a lower rate of sequence change compared with the rest of the genome; sequence relationships are indicated with an interrupted line to indicate this uncertainty. Three samples contained sequences distributed in more than one clade in V4 and V5, in the combinations tabulated in LN (B), CP (C), and LF (D) regions of the brain.

likely than the most likely tree calculated by using maximum likelihood (DNAML; $P = 0.018$).

In the V4/V5 region, sequence relationships were more complex and also discordant from those observed both for p17^{gag} and for V1/V2 or V3 (Fig. 2C, 2D, 3C, and 3D). Variants from the LN region formed three lineages (A, B, and C), with a fourth lineage comprising variants from the LF and RF regions (lineage D). However, approximately half of the LF variants grouped with lineage C. Similar sequence relationships were observed for the V5 region, with three main lineages comprising LN (A), LN and LF (B), and LF and RF (C). Because V4 and V5 sequences were amplified in the same PCR fragment, it was also possible to observe different combinations of lineages between the two regions (Fig. 3D). For example, V4 sequences of 12 LF clones were in lineage C in V4 and in lineage B in V5 (i.e., CB), while 9 clones were DC and 7 were CC. Similar reassortments were observed in LN sequences (one AA, two AB, one BB, and three BC).

To investigate further the tissue distribution of variants in different genomic regions, we obtained sequences from the *pol*, V1/V2, V3, and gp41/*nef* regions of variants recovered from other parts of the brain (Fig. 4). The most conserved region was the *pol* region, in which sequences from different regions of the brain did not display any significant phylogenetic groupings. The mean pairwise distance between *pol* sequences from different samples was 0.013 (0.048 at synonymous sites), lower than observed between variants within either lineage A or lineage B in the p17^{gag} region. Either there was a lower rate of sequence change at synonymous sites in this part of the genome or the *pol* region of the genome originated after the

diversification of p17^{gag}, implying the existence of recombination between the *pol* region and more variable regions of the genome.

In V1/V2, we observed four bootstrap-supported lineages, differing from each other by distances of 0.022 to 0.076, compared with distances of 0 to 0.008 within lineages. The distribution of sequences from different regions of brain did not match the distribution of sequences into the two lineages in p17^{gag} (Fig. 3 and 4). For example, sequences from LN (A in p17^{gag}) and choroid plexus (CP) (B in p17^{gag}) regions were found in the same lineage in V1/V2, while sequences from LF and RP regions, which grouped with the LN sequence in p17^{gag}, were on three separate lineages in V1/V2 (A, C, and D).

Similar complexity was observed among sequences from V3, V4, V5, and gp41/*nef*, with variants from the CP showing the greatest diversity in V3, V4, and V5. A subpopulation of CP variants grouped with those found in lymphoid tissue (e.g., lineages C in V3, B in V4 and V5, and A in gp41), but as described above for LN and LF, there were inconsistent relationships in the V4 and V5 regions (Fig. 3C), with all four combinations of lineages B and C in V4 and A and B in V5 being observed. Overall, no consistent relationship between lineages was observed among the other samples collected from the brain (Fig. 3), and apart from LO and RO regions, each region of the brain contained a different combination of phylogenetically distinct subgenomic fragments.

Tissue-specific grouping. Variants from different regions of the brain of NA234 consistently grouped separately from those recovered from LN in the V1/V2, V3, and gp41/*nef* sub-

A) V1/V2

124

216

A	CP	c	PLCVTLKCKD VGNITNITSS SWEKMEKGEI KNCSFKVITS IKDRTOIEHA LFKYKLDVVI DNDNDNTSYR LISCHTSVIT QACPKVSEF
B	RT	cR.T. .R.N..T... ..NI... .S..V.K... ..N.....
C	LF	11R.T. .R.N..T..N ..NI... .S..V.K... ..N..... .NK--.....
D	BS	cT. .R.N..T... ..NI... .S..V.Q... ..N.....
	LO	cTN .R.N..T... ..NI... .S..V.Q... ..N.....
	RO	cT. .R.N..T... ..NI... .S..V.Q... ..N.....
	RP	cT. .R.N..T... ..NI... .S..V.Q... ..N.....

B) V3

276

357

A	LN	1	VIRSENFNN AKIIIVHLNK SLAINCTRPN SNTRKGIHIA PGRAFYATGD IIGDIRQAHN NLSKAEWENT LKQVAMKIRE Q
B	LN	2T..... .V..H..... N..... ..I.R.D.N..... ..I..K..
C	LN	9T..Q.T. PVV..... N..... ..R.D.N..... ..I..K..
D	BS	cN..Q.T. .V..... N...S... ..A..T... ..Y..... ..I.....
	CP	2N..Q.T. .V..... N...S... ..T..T... ..R..I..D.N
	RP	cN..Q.T. .V..... N...S... ..T..T... ..R..I..D.N
	RT	cN..Q.T. .V..... N...S... ..T..T... ..R..I..D.N
	LF	16V.N..Q.T. .V..... N...S... ..T..T... ..I.....T..Q..I.....
	LF	12N..Q.T. .V..... N...S... ..T..T... ..R..I.....
	CP	3N..Q.T. .V..... N...S... ..A..T... ..I.VK.....
	RO	cN..Q..V..... N..... .A..T... ..I..K.....
	LO	cN..Q..V..... N..... .A..T... ..I..K.....

C) V4

392

445

A	LN	2	STQLFNSTWD NGEVL--SNT EGNQTTTLPK RIKQIINMWQ EVGKAMYAPP IR
B	LN	3P.GT--NI... ..DNI..... ..L.....
C	LN	4N.GT.GTT-N.. GE--NI..... ..V.....
	CP	7G..N.GT.GTT-N.. GE--NI.....
	BS	8N.GT.GTT-N.. GE--NI.....
	LF	19N.GT.GTT-N.. GE--NI.....
	RP	10N.GT.GTT-N.. GE--NI.....
	LT	10N.DT.GTT-N.. GE--NI.....
	RT	10N.GT.GTT-N.. GE--NI.....
D	LF	9N.DTDIKKLN... ..NI.....
	RO	10N.DTDIKKLN... ..NI.....
E	LO	6N.GT...--N.. .R.DNI.....
	RO	9N.GT...--N.. .R.DNI.....
F	BS	2N.DTDERTNI... ..NI.....

D) V5

446

478

A	LN	4	GQISCSSNIT GIMLTRDGGN TT-NETTEIF RPPGG
B	LN	5R..... .MI..... I N.RDNK.....
	CP	6R..... .MI..... I N.RDNK.....
	BS	10R..... .MI..... I N.RDNK.....
	LF	12R..... .MI..... I N.RDNK.....
	RP	10R..... .MI..... I N.RDNK.....
	LT	10R..... .MI..... I N.RDNK.....
	RT	10R..... .MI..... I N.RDNK.....
C	LF	16R..... .MI..... I SKS.NDS.V.....
	RP	11R..... .MI..... I SKS.NDS.V.....
D	LO	4R..... .MI..... I N.GDNE..V..V..
	RO	5R..... .MI..... I N.GDNE..V..V..

E) gp41/nef

840

870

51

A	LN2	c	RAFRVAVIHIP RRIRQGFERA LL*LC*DG* * MGNKWSKSSR SGWDTVRERM RRAEPAAEPA AAGVGAASRD LERHGAITSSN
B	LN1	c*..**..*..* ..D.....
C	CP	3*..**..*..* ..E.....
D	RP	cV..... ..*..**..*..* ..RR.PA..... ..D.....
	LF	cV..... ..*..**..*..* ..M.PA..... ..D.....
	RP	cV..... ..*..**..*..* ..T.PA..... ..D.....
	LT	cV..... ..*..**..*..* ..T.PA..... ..D.M.....
	CP	3G..... ..*..**..*..* ..T.SA..... ..D.....
E	RO	cL..... ..---E.W* ..G..... M.PA..... ..D.....
	LO	cL..... ..---E.W* ..G..... M.PA..... ..D.....
	RT	cL..... ..---E.W* ..G..... M.PA..... ..D.....
	BS	cL..... ..---E.W* ..G..... M.PA..... ..D.....

FIG. 4. Comparison of inferred amino acid sequences of variants from different regions of the brain of NA234 in V1/V2 (A), V3 (B), V4 (C), V5 (D), and gp41 (E) /nef, using LN variants as reference sequences. Horizontal lines divide primary bootstrap-supported ($\geq 75\%$ of data sets) phylogenetic groupings of nucleotide sequences. Each number in the third column indicates the number of clones used to create the indicated consensus sequence. c, consensus sequence obtained by direct sequencing of PCR product. Symbols: -, sequence identity with LN sequence; -, gap introduced to presence sequence alignment; *, termination codon. Sequences are numbered by their positions in the HIV_{LAI} gp120 (A to D) or nef (E) sequence. BS, brain stem.

TABLE 2. p17 ^{Gag} and V3 sequence divergence

Comparison	Study subject	Distance ^a					
		p17 ^{Gag}			V3		
		Mean	Error	Rank ^b	Mean	Error	Rank
Within brain	NA021	0.034	0.009	3	0.015	0.008	4
	NA173	0.035	0.010	2	0.017	0.008	3
	NA128	0.012	0.006	4	0.035	0.011	1
	NA234	0.042	0.011	1	0.028	0.010	2
Brain-lymphoid tissue	NA021	0.040	0.009	3	0.043	0.014	3
	NA173	0.040	0.011	2	0.031	0.011	4
	NA128	0.023	0.008	4	0.080	0.018	1
	NA234	0.057	0.013	1	0.072	0.018	2

^a Mean J-C distance between variants within brain or between brain and LN or spleen (excluding CP).
^b Rank of variability (1 = most variable).

genomic regions. To investigate further the differentiation between lymphoid and brain-derived variants, sequences from the p17 ^{Gag} and V3 regions were obtained from 8 to 12 brain regions from three other individuals (NA021, NA173, and NA128) and compared with those amplified from the corresponding LN or spleen samples (Table 2; Fig. 5 and 6). Se-

quences from each individual were monophyletic in both p17 ^{Gag} and V3 upon comparison with each other and with epidemiologically unlinked subtype B sequences (listed in Materials and Methods).

In each of the three study subjects, p17 ^{Gag} sequences of variants infecting different regions of the brain and lymphoid

A) NA173

		276							357	
A	LN2	VIRSENFTNN	AKVIIVQLNK	SVEINCTRPN	NNTRKGIHIG	PGRAPYRTGE	IIGNIRQAHC	DLSKAAWNTD	LKQIVIKLRE	Q
	LN3I.....	N..GT.....
	BG	N..R.....
	BS	N..R.....
	LF	N..RT.....
	RF	N..A.....
	LT	N..A.....
	RT	N..A.....
	RO	N..G.....
	RP	N..A.....

B) NA021

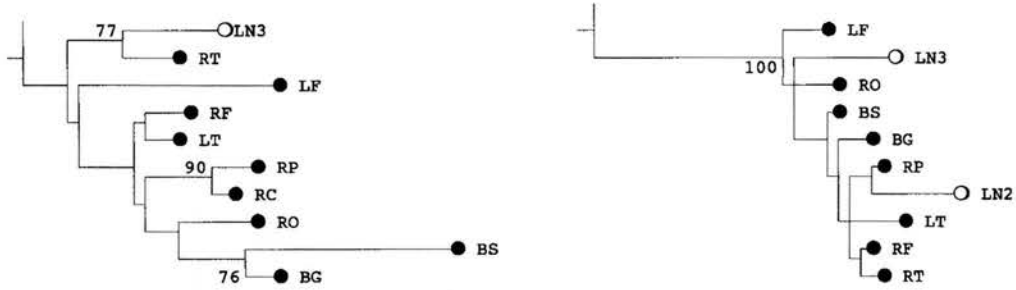
		276							357	
A	LN1	VIRSENFTNN	AKVIIVQLNE	TVEINCTRPN	NNTRKSITMG	PGRAPYATGD	IIGDIRQAHC	NISRTRWNNT	LKQIVIKLRE	Q
	LN2
B	LCN.....P..N.....
	RC
	BS
	LF
	RF
	RP
	BG
	CP
	RO
	RT
	LP
	LO

C) NA128

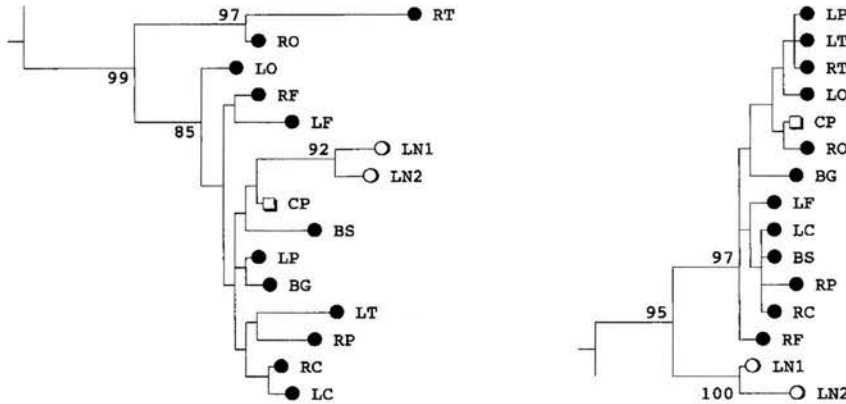
		276							357	
A	SPL	VIRSENFSDN	AKTIIVHLNE	SVEINCTRPN	NNTRKRITMG	PGRVYYTTCG	IIGNIRQAHC	NISRERWNKT	LEQIVIKLGE	K
	RF
	LF	T.....
B	RTN.....S..PI..AL.....
	LT
	LO
	RPN..V.....
	LP
	BG
	BS
	BS

FIG. 5. Comparison of inferred amino acid sequences of variants from the V3 region of NA173 (A), NA021 (B), and NA128 (C), using the LN or spleen (SPL) sequences as a reference. Horizontal lines divide bootstrap-supported ($\geq 75\%$ of data sets) phylogenetic groupings of nucleotide sequences (symbols and sequence numbering correspond to those in Fig. 4). BS, brain stem.

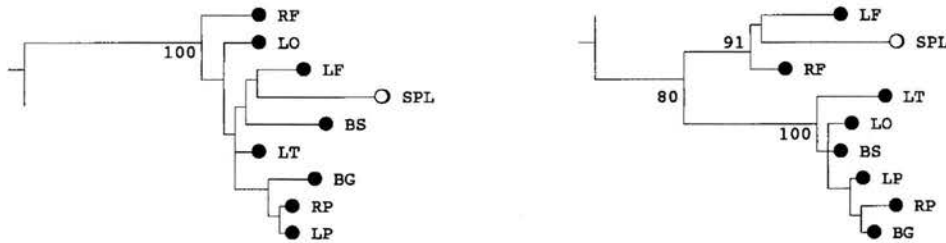
A) NA173



B) NA021



C) NA128



D) NA234

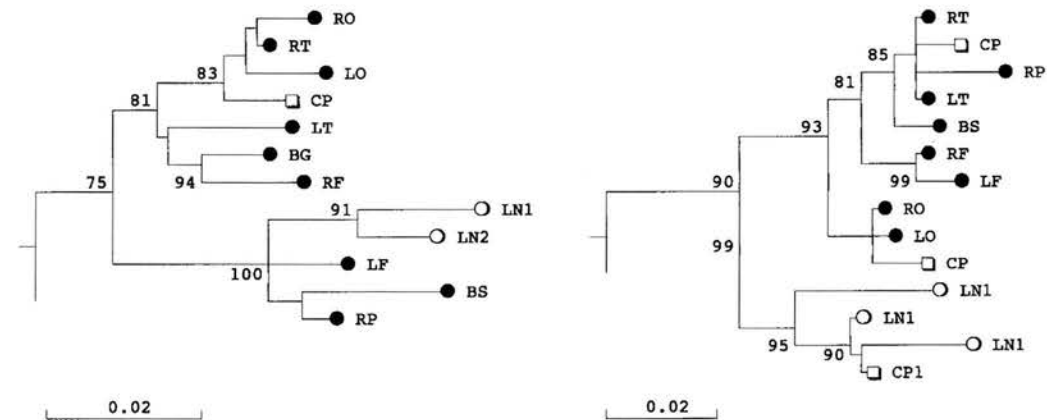


FIG. 6. Phylogenetic analysis of p17⁸⁹⁸ (column 1) and V3 sequences (column 2) of the four study subjects (A through D). Sequences from each of the four subjects were monophyletic in both genomic regions upon comparison with each other and the sequences available from GenBank listed in Materials and Methods and were used to root each clade. p17⁸⁹⁸ and V3 trees from different study subjects were plotted by using the indicated scale of J-C distances. Sequences were derived from brain (●), from lymphoid tissue (LN or spleen [SPL]) (○), or from CP (□). All bootstrap values of $\geq 75\%$ are indicated on branches. Sequences from NA234 were obtained from individual clones of amplified DNA instead of by direct sequencing; single representative clones from each sample or multiple clones for those containing sequences in more than one lineage (i.e., LN and CP) have therefore been included. BS, brain stem; BG, basal ganglia.

tissue showed interrelationships different from those observed in V3. First, there was no correlation between sequence diversity between the two genomic regions (Table 2). For example, p17^{gag} sequences from NA173 showed two- to threefold-greater variability between brain regions (mean J-C distance of 0.035 ± 0.010) and between brain and lymphoid tissue (mean J-C distance of 0.04 ± 0.011) than the corresponding sequences from NA128 (distances of 0.012 ± 0.006 and 0.023 ± 0.008). In contrast, sequences from NA128 in the V3 region were more variable than those from NA173 (0.035 ± 0.011 and 0.080 ± 0.018 , compared with mean distances of 0.017 ± 0.008 and 0.031 ± 0.011 for NA173).

Second, phylogenetic analysis of the p17^{gag} and V3 sequences produced different groupings of variants (Fig. 6). For example, LN sequences from NA021 clustered separately from brain-derived variants in the V3 region but grouped with brain variants in p17^{gag}. Conversely, LN sequences from NA173 were distinct from all but one of the variants in p17^{gag} but were undifferentiated from brain variants in V3. In each of the four study subjects, tissue specificity depended on the region being compared. A specific instance of incompatible phylogeny comparable to that observed for NA234 (see above) was observed in NA021, where LN sequences formed a bootstrap supported lineage in V3 distinct from the brain variants, while the primary branching order in p17^{gag} divided sequences from the RT and RO brain regions from the rest of the sequences.

Sequence divergence and pathology appearance. The severity of HIV-1 varied between the study subjects (Table 1), ranging from infrequent giant cells confined to perivascular regions (NA173) to widespread pathology affecting both white and grey matter (NA128; Fig. 7). The extent and type of HIV-induced pathology correlated with the degree of V3 but not p17^{gag} sequence diversity between different brain regions and in the extent to which V3 sequences from the brain differed from those in lymphoid tissue. V3 sequences from NA173 (who showed minimal HIV-related pathology) were largely undifferentiated from those detected in LN, with only an alanine-arginine change segregating by tissue (mean J-C distances listed in Table 2). At the other extreme, the spleen-derived sequence from NA128 had a predicted syncytium-inducing phenotype and differed at multiple sites from the nonsyncytium-inducing variants in the brain. In the p17^{gag} region, however, sequences from NA173 were the most variable, and those from NA128 were the least variable.

DISCUSSION

This study documents the extraordinary complexity of HIV populations *in vivo*, findings which have several implications in understanding mechanisms of tissue adaptation and the timing of entry of HIV into the central nervous system (CNS). The marked sequence diversity of the p17^{gag} region and the readily identifiable clusters of sequences in the hypervariable regions of *env* (V1/V2, V3, V4, and V5) of NA234 and other study subjects provided evidence for recombination between different regions of the genome. For example, variants infecting separate regions of the brain of NA234 were assembled from two different p17^{gag} lineages and a limited number of distinct hypervariable region lineages, often with different combinations within same autopsy sample (e.g., the V4 and V5 sequences in the LN, CP, and LF sequences [Fig. 3 to D]). In the other three study subjects, the different degrees of variability and the discordant phylogenetic groupings between p17^{gag} and V3 regions indicate a lack of genetic linkage between these two subgenomic regions and further support the hypothesis of frequent recombination *in vivo*.

To date, recombination has been most easily identified between different subtypes of HIV-1; for example, variants of HIV-1 from Thailand contain *gag* sequences resembling those of subtype A but distinct from subtype A in the *env* gene (7), while other viruses appear to have been generated by multiple recombination events (e.g., HIV-1_{MAL} [10, 37] and subtype I [21]). Recombination has also been observed upon infection with different strains of HIV-1, either experimentally in a chimpanzee exposed to the laboratory isolates HIV_{LAI} and HIV-1_{SF2} (47) or possibly through multiple exposure to two or more sources of HIV infection in a blood recipient and an injecting drug user (12, 49). We have now demonstrated that recombination also occurs within an infected individual between variants descended in each case from the original infecting strain. The finding that different parts of the HIV genome can have different evolutionary histories severely limits the concept of tissue specificity of variants of HIV *in vivo*, particularly if these conclusions are based on a single subgenomic region. For example, the brain-specific and frequently monophyletic nature of HIV sequences in the *pol* region (virodemes) of variants infecting antiviral agent-treated individuals (48) may not be reflected elsewhere in the genome. Indeed, sequence relationships in the *env* gene may differ substantially from *pol*, as variation in the former region is more likely to confer phenotypic differences in cellular tropism. The existence of recombination provides an explanation for discordant phylogenies between p17^{gag}, V1/V2, and V3 that we observed between brain (in the LF region) and LN sequences in three previous subjects (14, 23, 24). While sequences in V3 were tissue specific, sequences in the p17^{gag} and V1/V2 regions were diverse, and some evolutionary lineages were common to variants recovered from brain, lung, and LN. Our observations support the hypothesis that recombination may accompany the acquisition of antiviral resistance, as exemplified by the appearance of zidovudine-resistant mutants in the peripheral circulation which occurred without evidence for a comparable bottlenecking in *env* (6); V3 sequences showed no reduction in diversity during the process of population replacement in *pol* region.

The diversity of V3 sequences in different brain regions of the four study subjects was similar to a previous comparison of variants infecting different brain regions (mean pairwise distance between brain regions, 0.021 [8], excluding sequences from the LF region that were highly divergent in sequence and failed to group phylogenetically with sequences derived from other regions of the brain). The LF sequences may have originated from exogenous contamination of the PCR, or corresponded to an epidemiologically unlinked isolate in a case of mixed infection. In either case, the observed degree of sequence divergence was unlikely to have originated from sequence change over the course of infection within the study subject.

Greater degrees of sequence complexity may also originate from the presence of different infected cell types in a tissue sample. HIV sequences amplified from the choroid plexus of NA234 showed the greatest diversity in the *env* region, containing variants corresponding to those from lymphoid tissue and brain, consistent with the presence of virus from blood-derived cells and brain parenchyma. The proximity of these different cell types in the CP may provide an opportunity for recombination of HIV to occur, as well as a site of entry of HIV into the CNS. Without biological characterization of the variants found in the CP or elsewhere in the brain, it remains unclear whether recombinant genomes have been selected or represent random samplings of phenotypically identical viruses. However, the multiple recombination events observed in

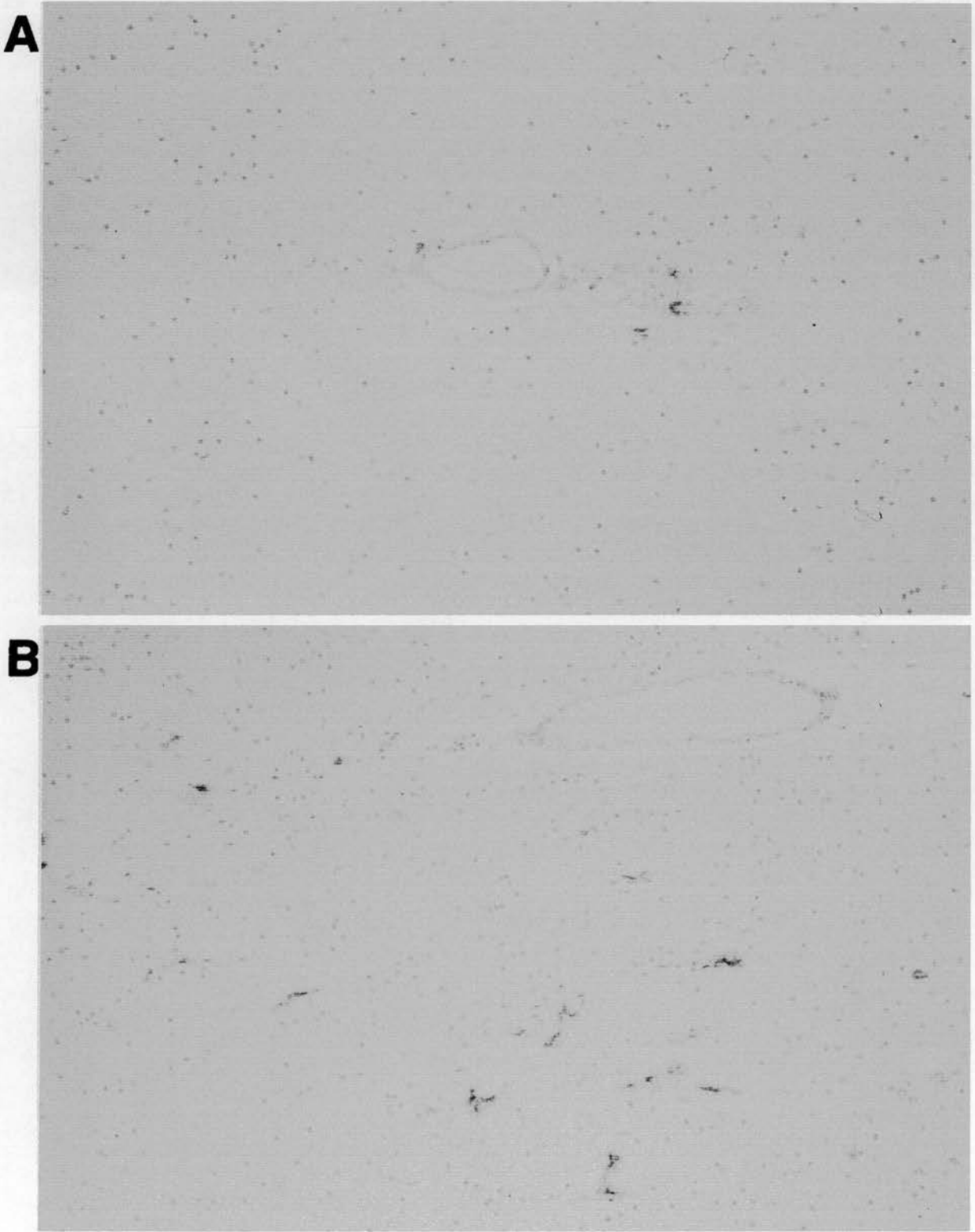


FIG. 7. Immunocytochemical detection of p24 antigen in representative sections of cerebral white matter of two study subjects with HIV. (A) NA173. Immunopositive mononuclear and giant cells (stained brown with diaminobenzidine) are confined to the perivascular region. (B) NA128. Widely dispersed HIV-infected microglia in white matter.

this study would provide a powerful mechanism for adaptation, providing, for example, an effective method for the spread of antiviral agent resistance or cytotoxic T-cell escape mutants into the CNS. Recombination between these latter determinants and *env* could produce new virus populations that retain their neuroadapted phenotype.

The differing sequence relationships between brain-derived and lymphoid variants from the four study subjects suggests that entry of HIV-1 into the CNS can occur at different times. The current consensus view that entry occurs early during HIV infection is supported by the observation that HIV RNA sequences can be detected in cerebrospinal fluid throughout the course of infection and by the detection of low levels of HIV proviral sequences in brains of asymptomatic individuals (4, 15, 41). Early entry is also supported by the extensive sequence diversity in the p17^{gag} region of variants recovered from the brain, such as between lineages A and B observed in NA234 in this study, which implies several years of divergent evolution (24). However, the relevance of early entry into the CNS in the development of late stage HIVE remains unclear, since active virus replication has not been demonstrated immunocytochemically during early infection (4), and brains show little evidence of pathology apart from the presence of infiltrating CD8 lymphocytes in perivascular areas.

Evidence for a contribution of late-entering variants to HIVE is provided by NA173, who showed a distinctive pathology appearance of HIV-expressing infiltrating macrophages confined to the perivascular regions (Table 1; Fig. 7A). The hypothesis of recent entry of HIV-infected cells into the brain parenchyma was supported by the observation of close sequence similarity in the V3 region of brain-derived variants with those obtained from lymphoid cells. This late-entry picture contrasted strongly with the distribution of HIV infection in NA128, in which HIV was widely dispersed in white (Fig. 7B) and grey matter, while V3 sequences were distinct between spleen and brain and heterogeneous within brain (Table 2; Fig. 5C). This correlation was, however, not supported by sequence comparisons in the p17^{gag} region, where sequence diversity was greatest in NA173 and least in NA128.

Indeed, to understand the adaptive significance of the sequence differences in different parts of the genome, it will in the future be necessary to analyze functionally the contribution of each genomic region to the phenotype of the virus. In particular, it will be important to determine the phenotypic significance of recombination between the p17^{gag} and *env* regions, particularly as variants with different combinations of lineages in the two regions were associated with distinct pathology appearances. Understanding what contributes to neurotropism will illuminate the selective pressures (if any) that produce the recombinant viruses observed in this study. The lack of genetic linkage in the HIV genome resulting from recombination greatly enhances its ability to adapt to several simultaneously acting selection pressures, as indicated by the rapid emergence *in vitro* of dual antiviral agent-resistant mutants (32).

ACKNOWLEDGMENTS

We are grateful to Francis Brannan for preparation of the sections for immunocytochemistry and for the provision of frozen samples from the MRC Edinburgh Brain Bank. We also greatly appreciate the critical review and discussion contributed by Donald Smith.

This study was supported by Medical Research Council Strategic Project grant G9632414.

REFERENCES

- Alkhatib, G., C. Combadiere, C. C. Broder, Y. Feng, P. E. Kennedy, P. M. Murphy, and E. A. Berger. 1996. CC CKR5: a RANTES, MIP-1 alpha,

- MIP-1 beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* 272:1955-1958.
- Asjo, B., J. Albert, A. Karlsson, L. Morfeldt Manson, G. Biberfeld, K. Lidman, and E. M. Fenyo. 1986. Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection. *Lancet* ii:660-662.
- Ball, J. K., E. C. Holmes, H. Whitwell, and U. Desselberger. 1994. Genomic variation of human immunodeficiency virus type 1 (HIV-1)—molecular analyses of HIV-1 in sequential blood samples and various organs obtained at autopsy. *J. Gen. Virol.* 75:867-879.
- Bell, J. E., A. Busutil, J. W. Ironside, S. Rebus, Y. K. Donaldson, P. Simmonds, and J. F. Peutherer. 1993. Human immunodeficiency virus and the brain—investigation of virus load and neuropathologic changes in pre-AIDS subjects. *J. Infect. Dis.* 168:818-824.
- Berson, J. F., D. Long, B. J. Doranz, J. Rucker, F. R. Jirik, and R. W. Doms. 1996. A seven-transmembrane domain receptor involved in fusion and entry of T-cell-tropic human immunodeficiency virus type 1 strains. *J. Virol.* 70:6288-6295.
- Brown, A. J. L., and A. Cleland. 1996. Independent evolution of the *env* and *pol* genes of HIV-1 during zidovudine therapy. *AIDS* 10:1067-1073.
- Carr, J. K., M. O. Salminen, C. Koch, D. Gotte, A. W. Arntstein, 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.
- Chang, J., R. Jozwiak, B. Wang, T. Ng, Y. C. Ge, W. Bolton, D. E. Dwyer, C. Randle, R. Osborn, A. L. Cunningham, and N. K. Saksena. 1998. Unique HIV type 1 V3 region sequences derived from six different regions of brain: region-specific evolution within host-determined quasiespecies—short communication. *AIDS Res. Hum. Retroviruses* 14:25-30.
- Cheng-Mayer, C., D. Seto, M. Tateno, and J. A. Levy. 1988. Biological features of HIV-1 that correlate with virulence in the host. *Science* 240:80-82.
- Cornelissen, M., G. Kampinga, F. Zorgdrager, and J. Goudsmit. 1996. Human immunodeficiency virus type 1 subtypes defined by *env* show high frequency of recombinant *gag* genes. *J. Virol.* 70:8209-8212.
- Deng, H. K., R. Liu, W. Ellmeier, S. Choe, D. Unutmaz, M. Burkhart, P. Dimarzio, S. Marmon, R. E. Sutton, C. M. Hill, C. B. Davis, S. C. Peiper, T. J. Schall, D. R. Littman, and N. R. Landau. 1996. Identification of a major co-receptor for primary isolates of HIV-1. *Nature* 381:661-666.
- Diaz, R. S., E. C. Sabino, A. Mayer, J. W. Mosley, and M. P. Busch. 1995. Dual human immunodeficiency virus type 1 infection and recombination in a dually exposed transfusion recipient. *J. Virol.* 69:3273-3281.
- Di Stefano, M., S. Wilt, F. Gray, M. Dubois-Dalcq, and F. Chiodi. 1996. HIV type 1 V3 sequences and the development of dementia during AIDS. *AIDS Res. Hum. Retroviruses* 12:471-476.
- Donaldson, Y. K., J. E. Bell, E. C. Holmes, E. S. Hughes, H. K. Brown, and P. Simmonds. 1994. *In vivo* distribution and cytopathology of variants of human immunodeficiency virus type 1 showing restricted sequence variability in the V3 loop. *J. Virol.* 68:5991-6005.
- Donaldson, Y. K., J. E. Bell, J. W. Ironside, R. P. Brettell, J. R. Robertson, A. Busutil, and P. Simmonds. 1994. Redistribution of HIV outside the lymphoid system with onset of AIDS. *Lancet* 343:382-385.
- Dragic, T., V. Litwin, G. P. Allaway, S. R. Martin, Y. X. Huang, K. A. Nagashima, C. Cayanan, P. J. Maddon, R. A. Koup, J. P. Moore, and W. A. Paxton. 1996. HIV-1 entry into CD4(+) cells is mediated by the chemokine receptor CC-CKR-5. *Nature* 381:667-673.
- Epstein, L. G., C. Kuiken, B. M. Blumberg, S. Hartman, L. R. Sharer, M. Clement, and J. Goudsmit. 1991. HIV-1 V3 domain variation in brain and spleen of children with AIDS: tissue-specific evolution within host-determined quasiespecies. *Virology* 180:583-590.
- Felsenstein, J. 1989. PHYLIP—phylogeny inference package (version 3.2). *Cladistics* 5:164-166.
- Feng, Y., C. C. Broder, P. E. Kennedy, and E. A. Berger. 1996. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* 272:872-877.
- Fenyo, E. M., L. Morfeldt Manson, F. Chiodi, B. Lind, A. von Gegerfelt, J. Albert, E. Olsson, and B. Asjo. 1988. Distinct replicative and cytopathic characteristics of human immunodeficiency virus isolates. *J. Virol.* 62:4414-4419.
- Gao, F., D. L. Robertson, C. D. Carruthers, Y. Y. Li, E. Bailes, L. G. Kostrikis, M. O. Salminen, F. Bibollettruche, M. Peeters, D. D. Ho, G. M. Shaw, P. M. Sharp, and B. H. Hahn. 1998. An isolate of human immunodeficiency virus type 1 originally classified as subtype I represents a complex mosaic comprising three different group M subtypes (A, G, and I). *J. Virol.* 72:10234-10241.
- Haggerty, S., and M. Stevenson. 1991. Predominance of distinct viral genotypes in brain and lymph node compartments of HIV-infected individuals. *Viral Immunol.* 4:123-131.
- Hughes, E. S., J. E. Bell, and P. Simmonds. 1997. Investigation of population diversity of human immunodeficiency virus type 1 *in vivo* by nucleotide sequencing and length polymorphism analysis of the V1/V2 hypervariable region of *env*. *J. Gen. Virol.* 78:2871-2882.

24. Hughes, E. S., J. E. Bell, and P. Simmonds. 1997. Investigation of the dynamics of the spread of human immunodeficiency virus to brain and other tissues by evolutionary analysis of sequences from the p17^{gag} and *env* genes. *J. Virol.* **71**:1272-1280.
25. Kasper, P., P. Simmonds, K. E. Schneeweis, R. Kaiser, B. Matz, J. Oldenburg, H. H. Brackmann, and E. C. Holmes. 1995. The genetic diversification of the HIV type 1 gag p17 gene in patients infected from a common source. *AIDS Res. Hum. Retroviruses* **11**:1197-1201.
26. Keys, B., J. Karis, B. Fadeel, A. Valentin, G. Norkrans, L. Hagberg, and F. Chioldi. 1993. 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**:475-483.
27. Korber, B. T. M., K. J. Kunstman, B. K. Patterson, M. Furtado, M. M. Mcevilley, R. Levy, and S. M. Wolinsky. 1994. 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.
28. Kozak, S. L., E. J. Platt, N. Madani, F. E. Ferro, K. Peden, and D. Kabat. 1997. CD4, CXCR-4, and CCR-5 dependencies for infections by primary and laboratory-adapted isolates of human immunodeficiency virus type 1. *J. Virol.* **71**:873-882.
29. Kumar, S., K. Tamura, and M. Nei. 1993. MEGA: Molecular Evolutionary Genetics Analysis, version 1.0. The Pennsylvania State University, University Park, Pa.
30. Li, Y. X., J. C. Kappes, J. A. Conway, R. W. Price, G. M. Shaw, and B. H. Hahn. 1991. 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**:3973-3985.
31. Liu, Z. Q., C. Wood, J. A. Levy, and C. Cheng Mayer. 1990. The viral envelope gene is involved in macrophage tropism of a human immunodeficiency virus type 1 strain isolated from brain tissue. *J. Virol.* **64**:6148-6153.
32. Moutouh, L., J. Corbeil, and D. D. Richman. 1996. Recombination leads to the rapid emergence of HIV-1 dually resistant mutants under selective drug pressure. *Proc. Natl. Acad. Sci. USA* **93**:6106-6111.
33. Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**:418-426.
34. Pang, S., H. V. Vinters, T. Akashi, W. A. O'Brien, and I. S. Chen. 1991. HIV-1 *env* sequence variation in brain tissue of patients with AIDS-related neurologic disease. *J. Acquir. Immune. Defic. Syndr.* **4**:1082-1092.
35. Power, C., J. C. McArthur, R. T. Johnson, D. E. Griffin, J. D. Glass, S. Perryman, and B. Chesebro. 1994. Demented and nondemented patients with AIDS differ in brain-derived human immunodeficiency virus type 1 envelope sequences. *J. Virol.* **68**:4643-4649.
36. Reddy, R. T., C. L. Achim, D. A. Sirko, S. Tehranchi, F. G. Kraus, F. Wong-Staal, C. A. Wiley, I. Grant, J. H. Atkinson, M. Kelly, J. L. Chandler, M. R. Wallace, J. A. McCutchan, S. A. Spector, L. Thal, R. K. Heaton, J. Hesselink, T. Jernigan, E. Masliah, I. Abramson, N. Butters, T. Patterson, S. Zisook, and D. Jeste. 1996. Sequence analysis of the V3 loop in brain and spleen of patients with HIV encephalitis. *AIDS Res. Hum. Retroviruses* **12**:477-482.
37. Robertson, D. L., P. M. Sharp, F. E. McCutchan, and B. H. Hahn. 1995. Recombination of HIV-1. *Nature* **374**:124-126.
38. Simmonds, P., P. Balfe, C. A. Ludlam, J. O. Bishop, and A. J. Leigh Brown. 1990. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. *J. Virol.* **64**:5840-5850.
39. 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.
40. Simmons, G., D. Wilkinson, J. D. Reeves, M. Dittmar, S. Beddows, J. Weber, G. Carnegie, U. Desselberger, P. W. Gray, R. A. Weiss, and P. R. Clapham. 1996. Primary, syncytium-inducing human immunodeficiency virus type 1 isolates are dual-tropic and most can use either Lestr or CCR5 as coreceptors for virus entry. *J. Virol.* **70**:8355-8360.
41. Sinclair, E., F. Gray, and F. Scaravilli. 1992. PCR detection of HIV proviral DNA in the brain of an asymptomatic HIV-positive patient. *J. Neurol.* **239**:469-471.
42. Steuler, H., B. Storch Hagenlocher, and B. Wildemann. 1992. Distinct populations of human immunodeficiency virus type 1 in blood and cerebrospinal fluid. *AIDS Res. Hum. Retroviruses* **8**:53-59.
- 42a. Strappe, P. Personal communication.
43. Stuyver, L., A. Wyseur, A. Rombout, T. Scarcez, C. Verhofstede, D. Rimland, R. F. Schinazi, and R. Rossau. 1997. Line probe assay for rapid detection of drug-selected mutations in the human immunodeficiency virus type 1 reverse transcriptase gene. *Antimicrob. Agents Chemother.* **41**:284-291.
44. Tersmette, M., J. M. Lange, R. E. de Goede, F. de Wolf, J. K. Eeftink Schattenkerk, P. T. Schellekens, R. A. Coutinho, J. G. Huisman, J. Goudsmit, and F. Miedema. 1989. Association between biological properties of human immunodeficiency virus variants and risk for AIDS and AIDS mortality. *Lancet* **i**:983-985.
45. Vanderhoek, L., C. J. A. Sol, J. Maas, V. V. Lukashov, C. L. Kuiken, and J. Goudsmit. 1998. Genetic differences between human immunodeficiency virus type 1 subpopulations in faeces and serum. *J. Gen. Virol.* **79**:259-267.
46. Vanderhoek, L., C. J. A. Sol, F. Snijders, J. F. W. Bartelsman, R. Boom, and J. Goudsmit. 1996. Human immunodeficiency virus type 1 RNA populations in faeces with higher homology to intestinal populations than to blood populations. *J. Gen. Virol.* **77**:2415-2425.
47. Wei, Q., and P. N. Fultz. 1998. Extensive diversification of human immunodeficiency virus type 1 subtype B strains during dual infection of a chimpanzee that progressed to AIDS. *J. Virol.* **72**:3005-3017.
48. Wong, J. K., C. C. Ignacio, F. Torriani, D. Havlir, N. J. S. Fitch, and D. D. Richman. 1997. In vivo compartmentalization of human immunodeficiency virus: evidence from the examination of *pol* sequences from autopsy tissues. *J. Virol.* **71**:2059-2071.
49. Zhu, T. F., N. Wang, A. Carr, S. Wolinsky, and D. D. Ho. 1995. Evidence for coinfection by multiple strains of human immunodeficiency virus type 1 subtype b in an acute seroconverter. *J. Virol.* **69**:1324-1327.