THE PATTERN AND PROCESS OF HUMAN IMMUNODEFICIENCY VIRUS SEQUENCE EVOLUTION *IN VIVO*

LIN QI ZHANG

PhD Thesis

Institute of Cell, Animal and Population Biology

University of Edinburgh

1992

DEDICATION

I would like to dedicate my thesis to mum, dad, sisters, and especially to my grandmother. It has been their sincere and persist love that have made everything possible.

献给我尊敬的妈妈，爸爸，姐姐和妹妹，特别是我的姥姥。是他们那无私和持久的爱一直在激励着我，我为我能有这样的家庭而感到骄傲自豪。

ACKNOWLEDGEMENTS

I owe the deepest gratitude to my supervisors, Drs. Andrew J. Leigh Brown and Peter Simmonds. It has been their advice and encouragement, as well as their own dedication to science, that profoundly stimulated my deep interest in virology and evolutionary biology.

I am indebted to Drs. Peter Balfe, Edward Holmes, Steve Burgess, Robert Walker, Christopher Ludlam and Henry Watson for their persistent help, patience and encouragement over the past four years. I would also like to express my deep appreciation to Mr. Sandy Cleland, Mrs. Fiona McOmish, Pamela Mackenzie, Liz Harvey, Selma Rebus, Sarah Ross, Ms. Sarah Ashelford, Yvonne Donaldson and Petra Zur Lage for their constant support and friendship throughout the course of study.

I wish to express my gratitude to all the patients who donated their samples and it has been their dedication and cooperation that really had made my project possible.

I would also like to give my greatest thanks to Sir Yue-Kong Pao, the Chinese government and British Council for their generous financial support during the past three years.

Finally, I would like to thank my parents, sisters and especially my grandmother. It has been their sincere and persistent love that carried me through all these years.

ABSTRACT

The research outlined in this thesis was primarily designed to study the quantitative and qualitative variability of plasma viral RNA during the course of HIV-1 infection. The quantitative aspect involved the development of a highly sensitive and reliable RNA-based PCR method which has been used to detect and quantify HIV RNA load directly from patient materials (plasma, serum). High levels of plasma viraemia (geometric mean value: 10^4 - 10^8 virion per ml of plasma) were observed during the primary stage of HIV-1 infection, considerably higher than those (geometric mean value: 10^3 - 10^4) observed in symptomatic patients. However, the high viral loads during this period are transient, and a marked drop in virus quantity was observed with the development of anti-HIV specific immune response. On average, HIV RNA was more abundant in the plasma of patients with more advance disease compared to asymptomatic patients. However, the observation of persistent high levels of HIV RNA in some asymptomatic patients suggests that viral replication continues throughout the course of HIV infection and that there is no 'latent' period to correspond with that observed with clinical progression.

Extensive studies of sequential sequence variation in the HIV-1 envelope gene constitute the qualitative element of this research and have revealed that there are complex evolutionary patterns. No sequence variation was observed in the V3 and V4 regions in any of the samples collected prior to or immediately after seroconversion, although variation was present in the *gag* gene at this time. Such an observation led to the suggestion that there is a strong selection for the most rapidly replicating viral

variants before the immune response is mounted. However, along with the development of specific anti-HIV immune response, the pattern and process of HIV-1 sequence variation changes. Rapid changes in the viral population were observed within weeks of seroconversion. Phylogenetic analysis of the V3 sequences from patient 82 identified several evolutionary lineages of virus variants after 3 years of infection, only two of which persisted and subsequently reached high frequency. Dramatic fluctuations in the population size of sequence variants were observed throughout the course of infection. Concurrently, selective constraints on the V3 region, and in particular of the V3 loop, were also evident as indicated by extensive convergent evolution (identical amino acid changes occurring in independent lineages). Thus, it seems likely that there are two major elements that are governing the evolutionary process of V3 sequences. One is the requirement for variability in order to facilitate 'escape' once virus variants are recognized by the immune system. The other is the constraint required to produce functionally viable viruses for further infection. The implication of these findings for our understanding of the biology of HIV are discussed in the thesis.

ABBREVIATION

ADCC	antibody-dependent cell mediated cytotoxicity
ag	attogram 10^{-18} gram
AIDS	acquired immunodeficiency syndrome
ARV	AIDS-related virus
ATP	adenosine-5'-triphosphate
AZT	3'-azido-3'deoxythymidine
bp	base pair(s)
BSA	bovine serum albumin
^{0}C	degrees centigrade
CAEV	caprine arthritis-encephalitis virus
Ci	Curie
CDC	Centers for Disease Control
cDNA	complementary deoxyribonucleic acid
CMV	cytomegalovirus
CSF	cerebrospinal fluid
CTL	cytotoxic T lymphocytes
dATP	deoxyadenosine-5'triphosphate
dCTP	deoxycytidine-5'-triphsphate
(d)dATP	2', 3'-dideoxyadenosine-5'-triphosphate
(d)dCTP	2', 3'-dideoxycytidine-5'-triphosphate
(d)dGTP	2', 3'-dideoxyguanosine-5'-triphosphate
ddC	2', 3'-dideoxycytidine
ddI	2', 3'-dideoxyinosine
(d)dTTP	2', 3'-dideoxythymidine-5'-triphosphate
DEPC	diethylpyrocarbonate
dGTP	deoxyguanosine-5'-triphsphate
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease

DNAML	Maximum likelihood
DTT	dithiothreitol
dTTP	deoxythymidine-5'-triphsphate
EDTA	ethlenediaminetetraacetic acid
EIAV	equine infectious anaemia virus
ELISAs	enzyme-linked immunosorbent assays
FCS	foetal calf serum
fg	femtogram 10^{-15} gram
FIV	feline immunodeficiency virus
GCG	Genetic Computer Group
HEPES	N-2-hydroxyethylpiperasin-N'-2-ethanesulphonic acid
HBV	hepatitis B virus
HCV	hepatitis C virus
HIV	human immunodeficiency virus
HLA	human leucocyte antigen
HTLV	human T lymphotropic virus
IDAV	immunodeficiency associated virus
Klenow	large fragment of DNA polymerase
LAV	lymphadenopathy-associated virus
mRNA	messenger ribonucleic acid
ng	nanogram (10^{-9} gram)
OD	optical density
^{32}P	β emitting isotope of phosphorous
PBMCs	peripheral blood mononuclear cells
PCP	pneumocystis carinii pneumonia
PCR	polymerase chain reaction
PBS	phosphate buffered saline
%	percentage
pg	picogram (10^{-12} gram)
PND	principal neutralization determinant
RNA	ribonucleic acid

RNase	ribonuclease
RT	reverse transcriptase
^{35}S	β emitting isotope of sulphur
SDS	sodium dodecyl sulphate
SIV	simian immunodeficiency virus
SSC	standard citrate saline
TCR	T cell receptor
TEMED	NNN'N'-tetra-methyl-1,2-diamino-ethane
tk	thymidine kinase
T*m*	melting temperature of double stranded nucleic acid
Tris	tris(hydroxymethyl)-amino-methane
UV	ultraviolet
μCi	microCurie
μl	microlitre
μM	micromolar
VV	visna virus
W	watt
w/v	weight per volume

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 GENERAL BACKGROUND: HIV AND AIDS

1.1.1 Natural History of HIV Infection

1.1.2 The Virus and Its Discovery

1.1.3 Biological and Biochemical Characteristics of HIV

1.1.4 Genomic Organization of HIV-1

1.1.5 The Structural Genes of HIV-1

1.1.6 The Regulatory Genes of HIV-1

1.1.7 Viral Genes of Uncertain Function

1.1.8 The Life Cycle of HIV-1

1.1.9 Latency and Persistence of HIV

1.1.10 Pathogenesis of HIV-1 Infection

1.1.11 Epidemiology and Transmission of HIV-1

1.1.12 Vaccine Development

1.2 EVOLUTION OF HIV

1.2.1 Evolutionary Origin of HIVs and SIVs

1.2.2 Extreme Variability of Lentiviruses Genome

1.2.3 Sequence Diversity in the *env* Gene of HIVs

1.2.4 Sequence Diversity in the *gag* and *pol* Genes of HIVs

1.3 PREVIOUS STUDIES OF SEQUENCE EVOLUTION IN THE EDINBURGH HAEMOPHILIAC COHORT

1.4 THE OUTLINE RESEARCH PRESENTED IN THIS THESIS

1.1 General Background: HIV and AIDS

The human immunodeficiency virus (HIV), the etiologic agent of the acquired immunodeficiency syndrome (AIDS), has the capability of selectively infecting and ultimately incapacitating the immune system. HIV induced profound immunosuppression results in a host defect that renders the body highly susceptible to opportunistic infection and malignant tumours associated with severe defects of cell-mediated immunity (Gottlieb *et al.*, 1981; Masur *et al.*, 1981; Nelson, 1990). The immune defect appears to be slowly progressive and irreversible. AIDS patients develop a range of infections such as pneumocystis carinii pneumonia (PCP), generalized cytomegalovirus (CMV) infection, progressive herpes and mucocutaneous candidiasis and tumours, particularly a tumour of the skin and viscera known as Kaposi's sarcoma and various types of lymphoma (Stahl *et al.*, 1982; Friedman-Kien *et al.*, 1982). HIV infection also causes at least 50% of AIDS patients to develop encephalopathy with loss of memory, impaired speech and dementia (Price *et al.*, 1986; Navia *et al.*, 1986*a, b*). AIDS was first recognized in the United States of America in 1981 (Gottlieb *et al.*, 1981) and subsequently the number of cases has increased rapidly. Up to the end of 1991, approximately 9-11 million people are thought to be infected worldwide and 1.5 million have so far developed AIDS (Chin *et al.*, 1991). The rapid spread of the AIDS virus not merely in homosexual men, but in haemophiliacs, intravenous drug abusers, heterosexual partners and children of infected parents indicates that the viral infection is transmitted through sexual contact, blood, blood products and perinatally (Chin *et al.*, 1991).

1.1.1 Natural History of HIV Infection Like other retroviral infections, HIV infections are chronic infections that probably persist for life (Clements *et al.*, 1988). Infectious virus has been successfully isolated from a majority of both asymptomatic and symptomatic seropositive individuals after seroconversion (Popovic *et al.*, 1984; Barre-Sinoussi *et al.*, 1983; Gallo *et al.*, 1984). After an initial acute illness, the patient may be asymptomatic for a prolonged period of time. Then, the infection may lead to gradual impairment of immune and neurological functions (Fauci, 1988). Several different staging classifications have been proposed to monitor the course of HIV infection. The Centers for Disease Control (CDC) and the Walter Reed (WR) classifications are those currently used most widely (Center for Disease Control, 1986; Redfield *et al.*, 1986). The patients studied in this work are all classified by CDC classification. The current CDC classification of HIV infections is depicted in Table 1.1

1.1.2 The Virus and Its Discovery First described as a novel disease in 1981 (Gottlieb *et al.*, 1981), AIDS was recognized several years before the new causative agent was identified and isolated. Several earlier candidates as the causative agent such as CMV (cytomegalovirus), HBV (hepatitis B virus), HTLV-I (human T-cell lymphotropic virus) were excluded when Francoise Barre-Sinoussi published the first report of a new virus from a patient with lymphadenopathy typical of some pre-AIDS cases (Barre-Sinoussi *et al.*, 1983). This new virus was classified in the subfamily *Lentivirinae* of the family *Retroviridae* (Barre-Sinoussi *et al.*, 1983; Gallo *et al.*, 1984). The family of Retroviridae consists of three subfamilies, *Oncovirinae,*

Table 1.1 The current CDC classification system for human HIV infection

---

Group I: Acute infection

Mononucleosis like syndrome, with or without aseptic meningitis, associated with HIV antibody seroconversion.

Group II: Asymptomatic infection

Possible subclassification:

A. No pathological laboratory findings;

B. Pathological laboratory findings.

Group III: Persistent generalized lymphadenopathy

Lymph node enlargement of 1 cm or greater at two or more extra-inguinal sites persisting for more than 3 months in the absence of a concurrent illness or condition other than HIV infection.

Group IV: Other disease

Subgroup A: constitutional disease (weight loss greater than 10%, fever longer than 1 month, diarrhoea longer than 1 month).

Subgroup B: neurological disease (dementia, myelopathy, peripheral neuropathy).

Subgroup C: secondary infectious disease (*pneumocystis carinii* pneumonia, chronic crytosporidiosis, toxoplasmosis etc.).

Subgroup D: secondary cancers (Kaposi's sarcoma, non-Hodgkin's lymphoma, primary lymphoma of the brain).

Subgroup E: other conditions

Spumavirinae, and *Lentivirinae* (Weiss *et al.,* 1985). Lentiviruses cause chronic disease affecting the lungs, joints, nervous, haematopoietic and immune systems of humans and animals (Weiss *et al.,* 1985). The lentivirus group are so called because the prototype virus of this subfamily, visna virus, can cause the 'slow' degeneration of the central nervous system in sheep (Haase, 1986*a, b*). The lentivirinae subfamily also contains other animal viruses such as EIAV (equine infectious anaemia virus), CAEV (caprine arthritis-encephalitis virus), SIV (simian immunodeficiency virus), FIV (feline immunodeficiency virus) and others (Clements *et al.,* 1988).

From the time this novel retrovirus was discovered, it has been variously named LAV (lymphadenopathy-associated virus), IDAV (immunodeficiency associated virus), HTLV-III (human T-cell lymphotropic virus type III) and ARV (AIDS-related virus) by different laboratories. But HIV (human immunodeficiency virus) has been adopted as the accepted nomenclature by the International Committee on the Taxonomy of Viruses in 1986 (Coffin *et al.,* 1986*a, b*). Not long after the first HIV was isolated, another novel isolate, which is genomically and antigenically related, but clearly distinct from the prototype HIV, was obtained from a western African AIDS patient (Clavel *et al.,* 1986*a, b*; Guyader *et al.,* 1987). It was therefore proposed to name the prototype isolates as HIV-1 and novel isolates as HIV-2. HIV-1 isolates comprise those viruses which are responsible for the current AIDS epidemic in Central Africa, Europe, the Americas and other regions of the world. However, the distribution of HIV-2, which has previously been identified as LAV-2 or HTLV-IV, is largely in West Africa.

The HIV virion is roughly 100nm in diameter. It has a double layer of lipid

which is derived from the outer membrane of the host cell. Studding the membrane are glycoproteins made up of two components: gp41 which spans the membrane and gp120 which extends beyond it. The membrane also contains other proteins such as HLA (Human Leucocyte Antigens) which are believed to be derived from the membrane of the human cell. This membrane-protein envelope covers a nucleocapsid core which is eikosahedral sphere in shape and made up of proteins designated as p24 (the capsid protein), p17(the myristoylated protein) and p9 and p7 (the nucleocapsid protein). The virus itself has two identical RNA molecules which are carried in the core along with several copies of reverse transcriptase (p66 and p51), protease (p22) and integrase (p32) (Gelderblom *et al.*, 1987). The schematic structure of HIV is presented in Figure 1.1.

1.1.3 Biological and Biochemical Characteristics of HIV After the association of HIV with AIDS was established, several biological characteristics of HIV were identified. Interestingly, HIV shares some biological and biochemical properties with Human T-lymphotropic viruses (HTLV-I and HTLV-II) and this is probably why HIV was initially called HTLV-III (Gallo *et al.*, 1984). The common features shared by HTLVs and HIV include a tropism for the helper T lymphocytes possessing the cell surface antigen CD4 (Dalgleish *et al.*, 1984; Klatzmann *et al.*, 1984); impairment of T-cell function (Barre-Sinoussi *et al.*, 1983, Popovic et al., 1984; Gallo *et al.*, 1984); induction of syncytia and formation of multinucleated giant cells in *in vitro* culture (Popovic *et al.*, 1984) ; absence of nucleic acid sequences derived from human DNA; and transcriptional regulation (*tat*) of viral and possibly cellular genes (Wong-Staal

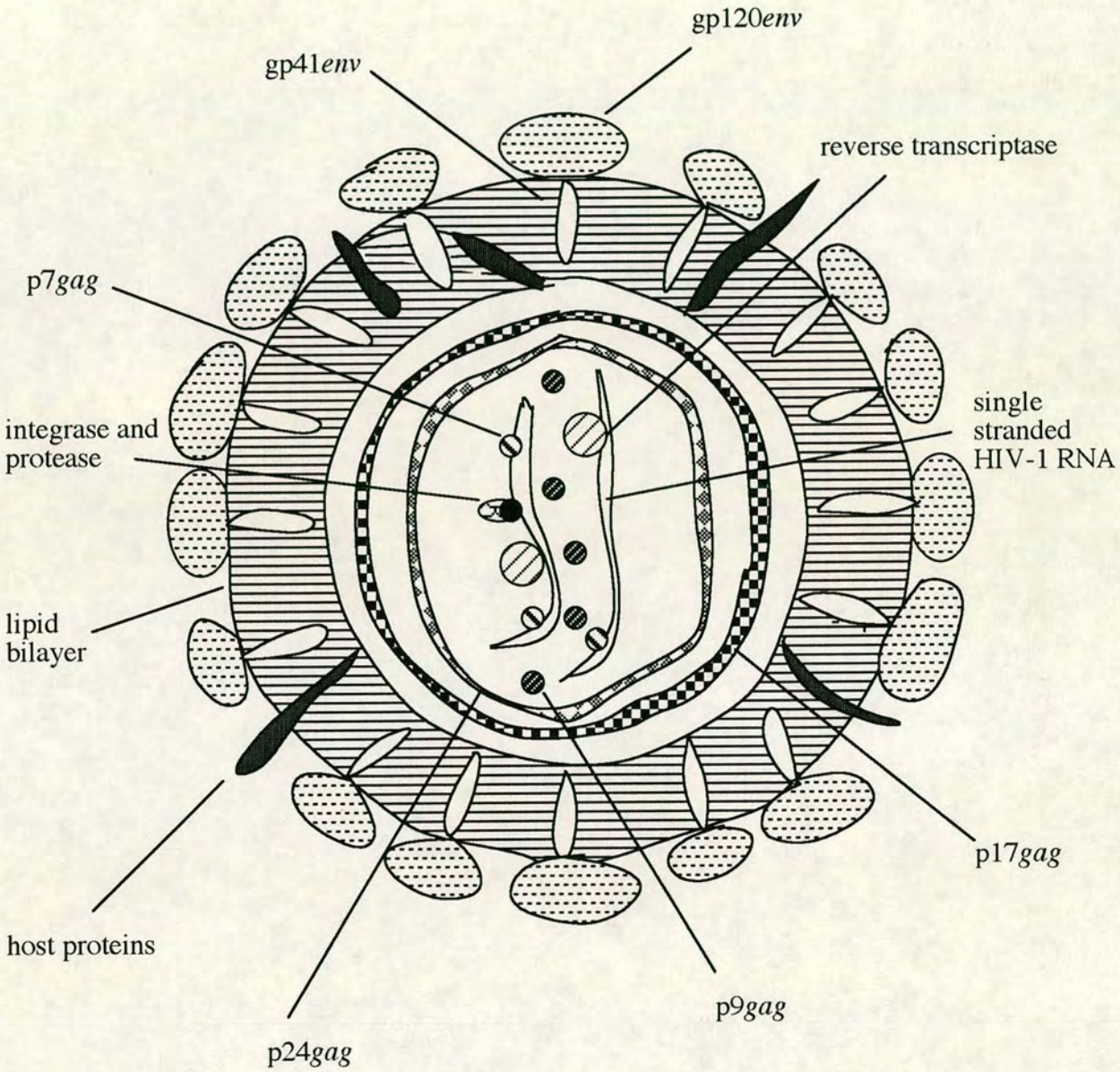
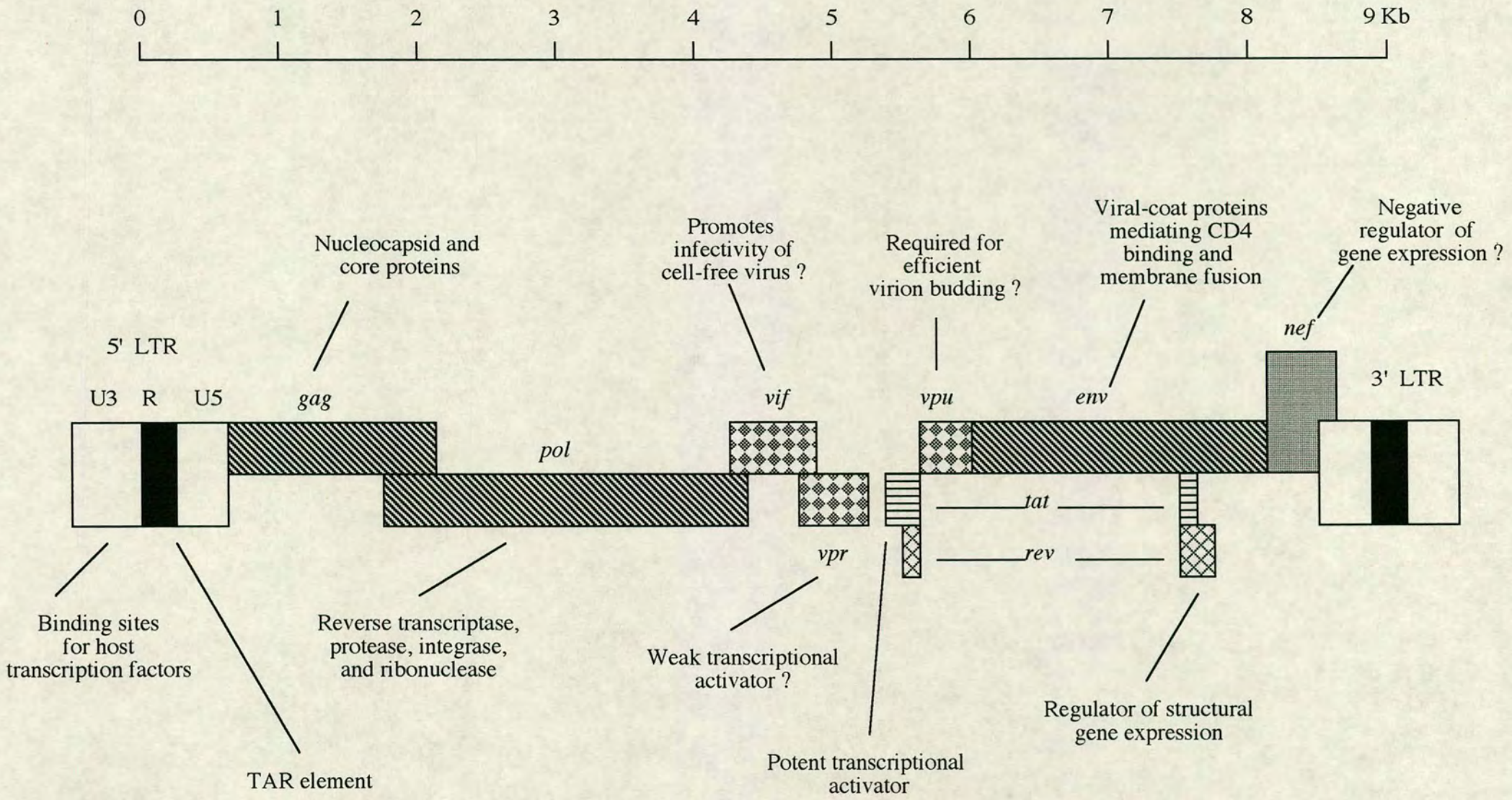


Figure 1.1. Schematic representation of HIV-1 virion

et al., 1985). However, unlike HTLVs, HIV shows cytopathic activity rather than immortalization and transformation of normal T lymphocytes (Chen *et al.,* 1983; Popovic *et al.,* 1984); extensive diversity in its genomic sequences probably relates to the highly replicative nature of the virus and the well-recognized infidelity of retroviral replication (Hahn *et al.,* 1984 and 1986; Willey *et al.,* 1986; Modrow *et al.,* 1987; Preston *et al.,* 1988; Roberts *et al.,* 1988; Coffin, 1992).

1.1.4 Genomic Organization of HIV-1 Schematic representation of the HIV-1 genomic organization is given in Fig 1.2. The genome of HIV-1 is about 10,000 nucleotides long in RNA form and comprises of at least nine genes (Ratner *et al.,* 1985; Wain-Hobson *et al.,* 1985). These genes are the *gag, pol* and *env* structural genes common to all replication-competent viruses, as well as genes involved in the regulation of viral replication (*tat, rev,* and *nef*), and genes of uncertain function (*vpu, vif* and *vpr*) (see Figure 1.2). These various gene products are translated from different mRNAs produced by the use of distinct splice donors and splice acceptors. In general, the viral regulatory proteins (*Tat, Rev* and *Nef*) are encoded by multiple spliced mRNA species while the structural and enzymatic viral proteins are the translated products of unspliced (*Gag, Pol*) or singly-spliced (*Env, Vif*) mRNAs (Cullen, 1991). Flanking these nine genes in its proviral DNA form are stretches of DNA called long terminal repeats, or LTR's which include DNA sequences that have important roles in controlling the expression of the viral genes (Sodroski *et al.,* 1984 and 1985*a, b*; Arya *et al.,* 1985).

Figure 1.2 Schematic representation of HIV-1 genomic organization. Function of each gene are shown and '?' indicates the uncertain function of corresponding gene.


0
1
2
3
4
5
6
7
8
9 Kb
5' LTR
U3
R
U5
gag
pol
vif
vpr
vpu
env
tat
rev
nef
3' LTR
Nucleocapsid and core proteins
Promotes infectivity of cell-free virus ?
Required for efficient virion budding ?
Viral-coat proteins mediating CD4 binding and membrane fusion
Negative regulator of gene expression ?
Binding sites for host transcription factors
TAR element
Reverse transcriptase, protease, integrase, and ribonuclease
Weak transcriptional activator ?
Potent transcriptional activator
Regulator of structural gene expression

1.1.5 The Structural Genes of HIV-1 The HIV-1 *gag* gene yields a *gag-pol* transcript which is subsequently translated to yield the 53-kd *Gag* precursor polypeptide. This precursor is then cleaved by viral protease to produce the p24 (phosphorylated), p17(myristylated), p9 and p7 *Gag* proteins (Veronese *et al.*, 1988). Together these polypeptides form the nucleocapsid of the HIV virion. The p7 protein contains a putative "zinc finger" domain that may be involved in direct interaction with the viral RNA (Veronese *et al.*, 1988). Like most myristylated proteins, p17 is associated with membrane structures and may serve to stabilize the exterior and interior components of the virion (Veronese *et al.*, 1988; Green, 1990).

The *pol* gene product of HIV-1 is translated from an unspliced *gag-pol* transcript by a mechanism that involves the frameshifting of the ribosome (Jacks *et al.*, 1987; Varmus, 1988*a, b*). The resultant *Pol* precursor protein is sequentially cleaved to yield the reverse transcriptase, protease, and integrase proteins (Varmus, 1988*a, b*). In addition to its polymerase activity, the HIV-1 reverse transcriptase contains an RNase H activity which is required for degradation of the RNA template during the synthesis of the double-strand DNA (Varmus, 1988*a, b*).

The HIV-1 *env* gene is transcribed as a single spliced viral mRNA species that when translated yields the 160-kd *Env* precursor protein. This precursor is subsequently cleaved and glycosylated in the endoplasmic reticulum and Golgi complex to yield the gp120 and gp41 glycoproteins (Stein *et al.*, 1990). While lacking a transmembrane domain of its own, gp120 is stabilized at the cell surface by its non-covalent interaction with gp41 (Veronese *et al.*, 1985; Helseth *et al.*, 1991). However, the release of gp120, which apparently occurs readily from the

virus or infected cells, has also been reported (Moore *et al.*, 1990).

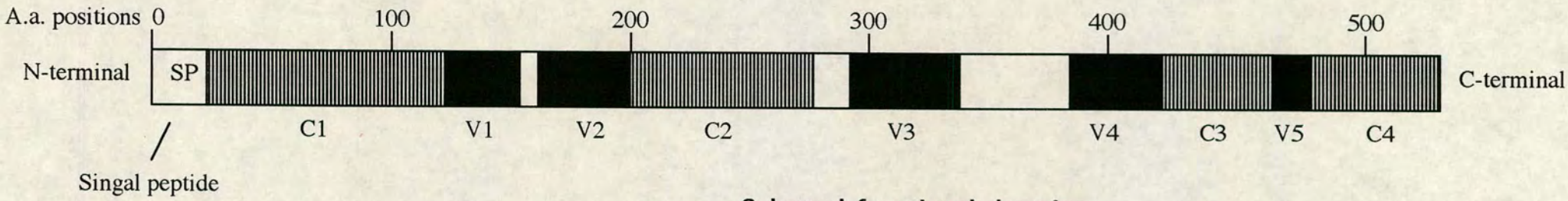
gp120 mediates attachment to the cellular virus receptor (CD4 molecule). The region of the ligand and the receptor involved in binding have been identified (Lasky *et al.*, 1987; Jameson *et al.*, 1988). By analyzing the crystallised N-terminal half of CD4 molecule secreted from Chinese hamster ovary (CHO) cells, the primary binding site for gp120 has been mapped in the first domain (D1) of the CD4 molecule with sequences resembling immunoglobulin variable (V) domains (Wang *et al.*, 1989; Ryu *et al.*, 1989). However, the precise three-dimensional structure of whole CD4 molecule has not yet been delineated, largely because of the poor diffraction of the whole CD4 molecule (Wang *et al.*, 1989). The location of other regions essential for virus infection and cytopathogenicity have been postulated, such as the region from amino acids 260-270 involved in the post-CD4 binding event (Bolognesi, 1990); the region from amino acids 303-337 involved in syncytium formation (Lifson *et al.*, 1986) and the region from amino acids 213-365 which contains the major determinant of T-cell and macrophage tropism (Cheng-Mayer *et al.*, 1990*a, b*; O'Brien *et al.*, 1990; Westervelt *et al.*, 1991 and 1992, Hwang *et al.*, 1991; Cann *et al.*, 1992).

Comparison of independent HIV-1 isolates has shown that the region of the envelope gene encoding gp120 is characterized by considerable sequence variation (Hahn *et al.*, 1986; Simmonds *et al.*, 1990*a*; Leigh Brown, 1991; Coffin, 1992). This variation is clustered into five major so-called hypervariable regions: V1 (amino acids 135 to 154), V2 (163 to 203), V3 (305 to 395), V4 (396 to 414) and V5 (459 to 469) (Modrow *et al.*, 1987). These five hypervariable regions are interspersed with highly conserved sequences and regions of intermediate variability: C1 (amino acids

38 to 134), C2 (204 to 304), C3 (415 to 458) and C4 (470 to 510) (Modrow *et al.*, 1987). The extensive sequence heterogeneity in gp120 complicates efforts to make an effective vaccine against the AIDS virus. Several studies have located the epitopes within gp120 immunogen that are responsible for generating neutralizing activity. One of these epitopes is within the third hypervariable domain, commonly referred to as the V3 loop which consists of a stretch of approximately 35 amino acids bound by two cysteine residues (Rusche *et al.*, 1988; Palker *et al.*, 1988; Javaherian *et al.*, 1989). Other regions of gp120 might also be involved in eliciting neutralizing antibodies, such as the amino terminal half (V1 or V2) (Ho *et al.*, 1992, McKeating *et al.*, 1992, personal communication), and/or the carboxyl terminal half (V4 or V5) (Haigwood *et al.*, 1990). Some recent experiments suggest that there are also conformational epitopes formed by different parts of gp120 (Profy *et al.*, 1990; Steimer *et al.*, 1991; Sattentau *et al.*, 1991; Ho *et al.*, 1991*a, b*). Interestingly, despite its high divergence, the 21 cysteine residues in the gp160 are completely conserved in all isolated reported (Tschachler *et al.*, 1990). Even more striking is that the conservation of these cysteine residues are shared by both envelope proteins of SIV and HIV-2 (Tschachler *et al.*, 1990), suggesting the indispensable role of these cysteine residues for viral envelope function. Schematic representation of the variable and constant regions of gp120 and its related function is summarized in Figure 1.3

The membrane spanning glycoprotein gp41 contains a fusogenic domain (the region which centrally involved in virus-cell and cell-cell fusion) which may located at its N-terminal end (Gallaher, 1987; Gonzalez-Scarano *et al.*, 1987), but other sites

Figure 1.3. Schematic representation of variable and constant regions of gp120 and its related functions. Amino acid positions are numbered according to HIV_{HXB2}. Variable and constant regins are drawed based on the work of Modrow *et al.* (1987). Viral neutralizing and CTL epitopes are presented according to Bolognesi (1990) and Cheng-Mayer (1990).



Selected functional domains

105-117	gp41 complementarity
124-445	major determinant of macrophage and T-cell tropism
269	post-binding, viral entry
303-337	post-binding, fusion
420-463	CD4-binding

Virus neutralization epitopes

120-200	epitopes located in V1 and V2 regions
290-340	Principle Neutralizing Determinant (PND)
390-510	epitopes located in V4 and V5 regions

T cell epitopes

290-340	CTL recognition in V3 region
410-440	CD4 binding + CTL recognition

on both gp120 and gp41 are likely to take part in the overall process of virus-cell fusion. The gp41 glycoprotein is therefore responsible for the cell fusion and syncytia formation and thus may contribute to the pronounced cytopathic effects of HIV-1. A number of studies also suggest that there are conserved regions in gp41 (amino acids 735-752) that can elicit neutralizing antibodies (Dalgleish *et al.*, 1988). In addition, there are also T-cell epitopes within both gp120 and gp41 associated with immune activity of cytotoxic T lymphocytes (Cease *et al.*, 1987; Takahashi *et al.*, 1988).

1.1.6 The Regulatory Genes of HIV-1 HIV-1 encodes a powerful transactivation protein termed *Tat* that dramatically increases the expression of all genes linked to the retroviral LTR's (Sodroski *et al.*, 1985*a, b*; Arya *et al.*, 1985). The 14-kd *Tat* polypeptide is translated from a double spliced viral mRNA species and is comprised of 86 amino acids. The *Tat* protein is primarily localized in the nuclei of expressing cells (Hauber *et al.*, 1987) and it functions in a sequence-specific manner. *Cis*-acting viral sequences required for *Tat* protein, designated the TAR (transactivation response element), are located between +1 and +60 in the viral RNA (Rosen *et al.*, 1985). Common to the 5' terminus of all HIV-1 mRNAs, this TAR region appears capable of forming an RNA stem-loop structure both *in vitro* and *in vivo* (Feng *et al.*, 1988). At present, it remains unknown whether the *Tat* protein directly binds to this RNA loop or alternatively modifies the RNA binding activity of host factors.

Like *tat*, the *rev* gene is also essential for viral replication. The product of the *rev* gene is a 20-kd protein which is primarily localized in the nucleus of the

expressing cells (Green, 1990). The HIV *Rev* protein appears to play a crucial role in promoting the transmission from early regulatory gene expression to late structural gene expression, by activating the transport of the large pool of unspliced or partial spliced viral mRNA from the cell nucleus to the cytoplasm (Sodroski *et al.*, 1986*b*; Feinberg *et al.*, 1986). *Rev* also functions in a sequence-specific manner acting through a *Rev* response element (RRE) located in the *env* gene (Malim *et al.*, 1989*a*, *b*).

The *nef* (negative factor) gene of HIV was originally recognized as an open reading frame located near the 3' end of the retroviral genome. This open reading frame is conserved in most strains of HIV-1, HIV-2 and SIVs, thus may play an important functional role. Nef is a 27-kd myristylated protein (Guy *et al.*, 1987) primarily localized in the cytoplasm of expressing cells associated with membranous structures due to its myristic acid anchor. The *Nef* protein may act as a transcriptional silencer to inhibit activation of HIV-1 LTR's via a negative regulatory element (NRE) (Niederman *et al.*, 1989). However, more recent studies have failed to confirm this finding and the real function of this conserved retroviral protein thus remains uncertain (Hammes *et al.*, 1989). The most recent data, however, suggests that the *Nef* protein is required for maintaining high virus loads during the course of persistent infection *in vivo* and the *Nef* is required for full pathogenic potential (Kestler *et al.*, 1991).

Apart from HIV-1 self-regulating factors, some cellular factors have been found to influence the level of viral expression through interaction with the U3 region of 5' LTR. NF-kappa B protein purified from Nalmalwa B cells (Kawakamin

et al., 1988), for example, is likely to interact with the kappa B enhancer which identified between nucleotides -104 and -81 of HIV-1 LTR (Rosen *et al.,* 1985). Deletion and site-directed mutation of this viral kappa B-like enhancer markedly impaired effects on HIV-1 LTR activation induced by NF-kappa B protein (Greene, 1990). Three consecutive Sp1 sites have also been found in LTR, locating between nucleotides -77 to -46 (Jones *et al.,* 1986). Each of these site has been shown to bind to Sp1 transcription factor and mutation of these factor binding sites leads to the substantial loss of HIV-1 LTR activity both *in vitro* and *in vivo* (Jones *et al.,* 1986).

1.1.7 Viral Genes of Uncertain Function The *vif* (viral infectivity factor) gene appears to play a role in virion morphogenesis but is not absolutely required for viral replication *in vitro* (Varmus, 1988; Green, 1990). Proviral clones containing mutations in the *vif* gene yield only low levels of infectious particles, although, cell to cell transmission of virus is still possible at reduced efficacy.

The *vpu* gene also appears to be involved in virion morphogenesis promoting efficient assembly and release of the virus (Cullen, 1991; Green, 1990).

The function of *the vpr* gene still remains unknown, although it has been suggested that it may play a role in accelerating virion assembly (Cullen, 1991; Green 1990).

1.1.8 The Life Cycle of HIV-1 Like that of all retroviruses, HIV-1's life cycle comprises a number of discrete steps, which include the binding of the virion to specific receptors on the surface of the target cell, penetration of the virus into the

cells, the synthesis of viral DNA, the integration of viral DNA into host cell DNA to form the provirus, the transcription of viral mRNA, and the assembly and release of mature virions (Varmus, 1988*a, b).*

The $CD4^+$ human T lymphocytes, B lymphocytes and macrophages form the primary cellular target for the infectious HIV-1 virion (Schnittman *et al.,* 1989; Dalgleish *et al.,* 1984; Klatzmann *et al.,* 1984). CD4+ cells are more easily infected by the HIV-1 virion in the solid tissues than they are in the blood circulation (Fauci, 1992; Meltzer *et al.,* 1990*a, b*). Recently, it has been reported that HIV-1 also appears able to infect $CD4^-$ cells such as neoplastic glial cells (Cheng-Mayor *et al.,* 1987; Watkin *et al.,* 1990), gut epithelial cells and a broad array of other nonlymphoid cells (Levy *et al.,* 1985; Cheng-Mayer *et al.,* 1987). The tropism of HIV-1 for CD4-expressing cells is readily explained by the finding that the membrane-bound CD4 protein forms the cellular receptor for HIV-1 (Dalgleish *et al.,* 1984). The external envelope protein, gp120, is believed to interact with the CD4 molecule (Lasky *et al.,* 1987). However, the CD4 independent entry of HIV-1 infection still remains a mystery. One possible explanation could be the presence of Fc receptors on some of the $CD4^-$ cells. Fc receptors (receptors for binding to the constant fragment [Fc] of an antibody which is responsible for binding the C1q component of complement) on the cell surface may therefore facilitate the combination with antibody-virus complex and consequently initiate the virus infection (Homsy *et al.,* 1989; Jouault *et al.,* 1989). *In vitro* experiment has indeed shown the initiation and enhancement of HIV-1 infection of fibroblast cells which are previously uninfectable before the expression of Fc receptors (McKeating *et al.,* 1990). Recent

experiments have shown that, like HIV-1, HIV-2 can also infect certain CD4- human cell lines (Clapham *et al.*, 1992; Zack *et al.*, 1990*a*), suggesting that HIVs can enter CD4- cells probably via an alternative cell surface receptor to CD4.

Subsequent to receptor binding at the cell surface, the HIV-1 virion is internalized by a process involving either cell fusion (Stein *et al.*, 1987) or receptor-mediated endocytosis (Maddon *et al.*, 1986). In the infected cell, viral DNA synthesis occurs in the cytoplasm within the nucleoprotein (Bowerman *et al.*, 1989). By using its own reverse transcriptase, a full-length linear double-stranded proviral DNA containing a single copy of all sequences present in viral RNA, plus duplication of the U3 and U5, is synthesised, which appears in the nucleus. At present, there is no information about the mechanisms responsible form nuclear transport or about the changes that occur in the nucleoprotein complex during or after migration to the nucleus. The final arrangement of linear DNA, U3-R-U5 --- viral genes --- U3-R-U5, is collinear with proviral DNA, which later on integrated into host genome by viral integrase, but includes two nucleotides at each terminus (TT at the 5' ends of each strand) that are removed during integration (Varmus, 1988*a,b*). Recent studies in a number of retroviral systems have provided convincing evidence that the linear unintegrated form of viral DNA present in the nucleus is the direct precursor to the integrated provirus ((Brown *et al.*, 1989; Fujiwara *et al.*, 1988). However, there are at least two forms of circular viral DNA have been found in the nucleus, the most abundant form contains a single copy of the LTR whereas smaller portion have two LTRs (Farnet *et al.*, 1991). Recent experiment carried out by Farnet *et al.*, suggested that circularization and integration are likely two independent and competing fates

form the newly formed linear viral DNA. In other words, the linear viral DNA will either be stabilised by integration into the host genome or forms circular molecules which no longer have the capacity to integrate.

Viral RNA is reformed by transcriptional activation of the HIV-1 provirus mediated through its 5' long terminal repeats (LTR's). Initiation of this process appears to require an "activated" T-cell environment and the presence of various inducible host transcription factors (Zack *et al.*, 1990 and 1992; Stevenson *et al.*, 1990; Bukrinsky *et al.*, 1991). Virus-encoded *trans*-acting factors (*tat, rev*) and several cellular factors with *cis*-acting sequences are required to control the HIV-1 gene expression. Initial products of HIV-1 gene expression are the short, multiple spliced RNA species encode *tat, rev* and, possibly *nef* (Cann *et al.*, 1989). The longer and full-length transcripts, which act as both virion RNAs and the mRNAs for the *gag-pol* polyprotein and for the envelope protein, then appear and accumulate in the cytoplasm. These findings demonstrate that there is a shift of HIV-1 gene expression from multiple-spliced short RNA species which encode regulatory protein to single- or un-spliced long RNA molecule, which largely encode viral structure proteins.

Following synthesis of viral proteins in the cytoplasm, several post-translation events occur. The envelope glycoprotein gp160 is cleaved and glycosylated in the endoplasmic reticulum and Golgi complex to yield the gp120 and gp41 glycoproteins (Stein *et al.*, 1990). Other proteins are either phosphorylated such as p24 or myristylated such as p17 and *nef* before they involved in the virus assembly. The process of viral assembly, maturation into infectious particles and release from the

cell via the process of budding is poorly understood. There seems a packing signal in the *gag* region which contribute to virus packaging. The sequence appears capable of forming a hairpin loop structure and may act as a primary sequence for binding of *gag* proteins (Cann *et al.,* 1989). Finally, somehow assembled mature virions bud from the cellular membrane to promote dissemination of the virus to other $CD4^+$ and probably CD4- cells, and at the same time, may result in the death of the original infected cells (Popovic *et al.,* 1984;).

1.1.9 Latency and Persistence of HIV Existing in their latent stages is probably one of the major means by which lentiviruses persist in their hosts. According to our present knowledge, there are at least three potential forms of latent HIV-1. First, resting T cells appear deficient in various cellular factors required for the efficient conversion of retroviral RNA genome into a double-stranded DNA equivalent (Fritsch *et al.,* 1977; Varmus *et al.,* 1977). Thus, one form of latent virus may correspond to the accumulation of viral RNA and reverse transcriptase complex in the resting T cells. Alternatively, these RNA in the complex may be reverse transcribed, yielding complete or partial linear DNA duplex (Green *et al.,* 1990; Zack *et al.,* 1990*b* and 1992; Stevenson *et al.,* 1990). This viral DNA is normally incorporated into the host genome under the control of the viral integrase protein. However, this integration process appears to be dependent upon certain host factors lacking in resting T cells. Thus, this unintegrated DNA form of HIV represents a second form of latent virus, although it is not stable in long term. A third form of latent HIV-1 may exist at the level of the fully integrated provirus. In the normal course of the immune response,

some of the activated T cells eventually return to the G_0 phase of the cell cycle as memory T cells. Because of the unlimited life span of memory T cells, HIV may persist within the genome of these memory T cells as an untranscribed provirus for a long period of time. Such viral latency may contribute to the prolonged asymptomatic period observed clinically. Taken together, latent forms of HIV-1 may theoretically exist at the RNA, unintegrated DNA, and integrated DNA forms.

Infected monocytes and macrophages also contribute to the persistence of HIV. Generally speaking, monocytes and macrophages are relatively resistant to the cytolytic effect of HIV (McElrath *et al.*, 1989; Nicholson *et al.*, 1986; Harper *et al.*, 1986). In addition, the infected monocyte or macrophage in the central nervous system may be protected from certain immune effector cells (Koenig *et al.*, 1986; Joly *et al.*, 1991). These properties may allow monocytes and macrophages to serve as an important reservoir of HIV . Furthermore, a small percentage of infected and activated $CD4^+$ T lymphocytes can also survive HIV infection (Schnittman *et al.*, 1989) and further contribute to virus persistence.

Finally, the high degree of sequence variation especially in its envelope gene may also help HIV-1 to persist in the face of immune recognition and clearance. It has been suggested that the nucleotide substitution in the V3 region could alter the virus antigenicity and therefore deter neutralisation by autologous sera (Albert *et al.*, 1990; Wolfs *et al.*, 1991; Zwart *et al.*, 1991; Goudsmit *et al.*, 1991; Montefiori *et al.*, 1991).

1.1.10 Pathogenesis of HIV-1 Infection The impact of HIV infection on the immune system is manifested by marked changes in several aspects of immunological responses. However, the most profound immunosuppression in AIDS patients is mainly due to the depletion of $CD4^+$ T helper/inducer lymphocytes (Gallo *et al.*, 1984; Fauci, 1988; Lifson *et al.*, 1989). In normal individuals, CD4+ cells constitute approximately 50-70% of circulating T cells in blood, and play a vital role in the recognition of antigens and B cell activation (Lifson *et al.*, 1989). With HIV infection and replication, however, the $CD4^+$ T cells are killed by as yet unknown mechanisms. Recent experiments suggest that the HIV envelope glycoprotein may play an important part in killing $CD4^+$ T cells, probably through cell to cell fusion, although the precise mechanism for this cytopathic effect is unclear (Lifson *et al.*, 1986*a, b*; Sodroski *et al.*, 1986*a*). Additional mechanisms of $CD4^+$ T cell depletion may be due to an autoimmune response. $CD4^+$ T cells, infected or not, may be coated with free gp120 glycoprotein shed from the virus particle which are susceptible to be killed via antibody-dependent cell mediated cytotoxic (ADCC) mechanisms (Moore, *et al.*, 1990; Lyerly *et al.*, 1987*a, b*). This, or a similar mechanism, may also involve bone marrow precursor cells (Donahue *et al.*, 1987). However, all these hypotheses lack direct *in vivo* experimental evidence, and therefore depletion of $CD4^+$ T cells in AIDS patients still remains a mystery.

Recent experimental evidence from in vivo samples supports the autoimmune hypothesis. It was discovered that there is a selective depletion of T cells in AIDS patients that have specific T cell receptor V_β sequences, whereas V_α usage was normal (Imberti *et al.*, 1991). Normally, most antigens are recognized through their

interaction with the variable portion of the T cell receptor (TCR) α and β chains. However, T cells recognize another category of ligands, the superantigens, on the basis of the expressed V_β region alone, independently from the other variable TCR segment (Fleischer *et al.*, 1991*a, b*; Hugin *et al.*, 1991). The pattern observed by Imberti *et al.*, was compatible with superantigen action. Therefore, it has been proposed that HIV can encode a kind of "superantigen" which initially produces a massive stimulation of immune cell activities, but ultimately they lead to the cells' dysfunction and death.

$CD4^+$ T cells are a central component of the immune response and intimately involved with monocytes and macrophages, cytotoxic T cells, natural killer cells and B cells. Therefore, even a selective depletion of the $CD4^+$ T cell population can result in a multitude of immunologic deficits, leading to the life-threatening opportunistic infections characteristic of AIDS.

Apart from $CD4^+$ T cells, many other cells can also be infected and affected by HIV. The elimination of these non-T cells or an alteration in their function can also create immune abnormalities. Recent studies have shown that monocytes and macrophages can be infected by HIV, especially those in certain bodily tissues such as the central nervous system, lymph nodes, and lungs (Ho *et al.*, 1985; Meltzer *et al.*, 1990*a*; Gendelman *et al.*, 1989). The frequency of HIV-infected monocytes and macrophages in these tissues may be many time higher than that in the blood, and this may suggest that HIV infection may account for the dysfunction of the mononuclear phagocyte system (Meltzer *et al.*, 1990*a*). An additional consequence of frequent infection of macrophages in the central nervous system and in

cerebrospinal fluid (CSF) is to cause dementia complex in AIDS patients (Price *et al.*, 1986; Navia *et al.*, 1986*a, b*). Therefore, taken together with the similarities between HIV and the other lentiviruses that can infect macrophages in the central nervous system and induce encephalitis, it strongly suggests that HIV is neurotropic and that the central nervous system, at the same time, may serve as a sanctuary site for the virus.

HIV can also infect B lymphocytes. B-cell abnormalities, consisting of polyclonal activation with high immunoglobulin levels (Pahwa *et al.*, 1985) and a poor antibody response to novel antigens (Mizuma *et al.*, 1988), are common in AIDS patients, and may be the direct consequence of HIV infection. In addition, this heightened production of nonspecific immunoglobulin may also result in autoimmune processes (Lyerly *et al.*, 1987*a, b*; Edelman and Zolla Pazner, 1989).

1.1.11 Epidemiology and Transmission of HIV-1 A decade have passed since the first reported cases of AIDS were announced in the United States (Gottlieb *et al.*, 1981). The World Health Organization estimates that during the last 10-12 years, approximately 9-11 million people have been infected with HIV (5-6 million in men and 4-5 million in women) and of which 1.5 million have developed AIDS (Chin, 1991). The world-wide distribution of reported AIDS cases is not even, with particular high prevalence in America 277,042, Africa 151,455 and Europe 66,783 (World Health Organization global Statistics, 1992). However, the high infection rates of HIV in Asia (particularly in Thailand and India) and Latin America serve a warning sign for the potential rapid dissemination of virus in the rest part of the

world. In 1988, for instance, there was an explosive increase in HIV prevalence among injecting drug users (IDU) in Thailand and in 1989, the subsequent spread of HIV has been found in male and female prostitutes, sexually-active heterosexual men, and blood donor (Weniger *et al.,* 1991). In contrast, the rate of HIV infection in North America and West Europe remains stable in the past few years, probably due to the changes in behaviour (safer injection, protected sex, etc) (Chin, 1991).

In North America and Europe, the overall distribution pattern of HIV prevalence in various risk groups seems consistent, with highest frequency in homosexual/bisexual males which constitute nearly 65% of reported AIDS cases in the United States (Center for Disease Control Statistics, 1992). IDUs, on the other hand, have consistently constituted a second largest transmission category (19%) (Center for Disease Control Statistics, 1992). However, within countries or different geographic continents, the prevalence of HIV in different risk groups varies. The dramatic spread of HIV infection among IDUs in Bangkok in early 1988 and in South-east Asia represents the major proportion (31%) of HIV-positive individuals in that area (Weniger *et al.,* 1991). HIV infection through heterosexual contact constitutes almost 60% of reported AIDS cases in Thailand by the end of year 1990 (Weniger *et al.,* 1991), and may also account for the vast majorities of AIDS cases found in Central Africa (Chin, 1991).

In United Kingdom, the overall distribution pattern of HIV-1 prevalence in various risk groups is fairly consistent with that obtained from the United States, with 60% in homosexual men, 13% in IDUs, and 7% in haemophiliacs (AIDS News Supplement, CDS Weekly Report, 1992). However, the situation observed in

Scotland may represent a special case of HIV transmission in this country. By the end of June 1992, approximately 1,813 individuals have been found to be HIV-1 positive in Scotland and 902 of them (49.8%) are IDUs (AIDS News Supplement, CDS Weekly Report, 1992). HIV infection in Edinburgh is the worst of all among Scottish cities. About 53% of all Scottish HIV infections are found in Edinburgh where in Glasgow even with double amount of human population, HIV-infected individuals consist only 25% of Scottish HIV-positive patients. Furthermore, 59% of all IDUs infections in Scotland are found in Edinburgh and this has always been the situation from the initial outbreak of HIV infection in IDUs in early 80's in Scotland (AIDS News Supplement, CDS Weekly Report, 1992).

Whereas transmission of HIV infection in North America and Europe has been predominately among homosexual/bisexual men and IDUs, the predominant risk behaviour worldwide has been heterosexual activity (Chin, 1990 and 1991). In 1984, all of the reported AIDS cases were in homosexual men, but since then there has been a gradual increase in the percentage of reported heterosexual AIDS cases to the extent that by 1986/1987 most reported AIDS cases were in heterosexuals (Chin, 1990). These data clearly indicate that the incidence of HIV infection among heterosexuals began to increase markedly during the early to mid-1980s (Chin, 1990). There is no doubt that high incidence of HIV infection in heterosexuals will also raise the potential for transmission of HIV infection from mother to child, either *in utero*, perinatally or postnatally. The relative contribution of each of these routes of transmission remains unknown. Published estimates of vertical transmission rates vary between 7 to 39% of birth to HIV-infected mothers (Newell *et al.*, 1990). The

estimated rates of vertical transmission from the Europe Collaborative Study (13%), based on information from 10 European centres, is markedly different from the rates reported from Africa (33% in Rwanda and 39% in Zaire and Zambia) (Newell *et al.*, 1990). The reason for these differences are not clear. There is little information on maternal factors that influence vertical transmission. Mother's clinical status during pregnancy and the duration of her infection may be important, but evidence remains circumstantial (Newell *et al.*, 1990).

1.1.12 Vaccine Development Vaccination has proved a highly successful strategy in the control of a range of disease caused by viruses, including smallpox, yellow fever and poliomyelitis, and is one of the obvious routes in attempting to control AIDS. However, the development of a safe and effective vaccine for infection with HIV is complicated by several unique scientific, logistic and ethical issues. The first issue is the failure to delineate the viral components or epitopes that induce protective immunity in the host. Despite the enormous amount of research on the immunopathogenesis of HIV infection and the characterization of the immune response to the virus, the definition of protective immunity against initial infection with HIV remains unknown. The lack of an adequate and convenient animal model for studying HIV infection and disease is another issue which seriously limits the development of an effective HIV vaccine. Although chimpanzees can be readily infected by HIV, however, no disease has occur in these animals. Thus, these species can only be used as a model for developing a vaccine that prevents initial infection by HIV. In addition, the logistic constraints of the chimpanzee model, including the

availability of animals and the cost of care, make this model less than ideal. Most recent development in animal model studies have come in the SIV system. The progress have been made in several areas including 1) the first evidence of vaccine cross-protection induced against a divergent isolate; 2) the demonstration that serum plasma from infected or vaccinated animal can confer protection when administered passively; 3) the first evidence that recombinant-based vaccines can be efficacious (Marc *et al.*, 1991; Gardner *et al.*, 1991).

In the light of experience with existing virus vaccines and understanding of the process of viral infection, several possible approaches to the development of vaccines against HIV have been considered. These approaches include live attenuated vaccines, whole killed vaccines and subunit vaccines consisting of a single viral polypeptide which may be obtained either from the virus directly or by recombinant DNA technology. In recent years novel approaches to the development of vaccines have been suggested as a consequence of developments in recombination DNA technology, chemical synthesis and immunology. These new strategies include genetically engineered live virus vectors, chemically defined synthetic peptides and anti-idiotypes. However, the very nature of HIV infection raises objections to using live attenuated or inactivated virus vaccines, as it is difficult to ensure that they are non-infectious or completely inactivated. The most promising types of vaccine will therefore be those involved in expression of viral antigens in acceptable vectors either as virus vector, or recombinant DNA product, and in synthetic peptides. Again, with this strategy, the problem of identifying the antigen which can induce adequate protective immunity remains.

With the advent of modern genetic engineering technology, genetic manipulated viruses and bacteria have improved the repertoire of antigen delivery systems available for vaccine development. Live recombinant viral vector have a number of attractive features, including the potential to resemble a live-virus-like infection and the ability to stimulate both humoral and cell-mediated immunity. Live recombinant vaccinia virus vectors were the first recombinant vectors to be developed, and a large number of different viral, bacterial and parasitic antigens have been expressed in this system. Recombinants expressing a variety of different HIV and SIV genes have been constructed and tested from immunogenicity in animal, and in some cases, man (Earl *et al.,* 1991; Shen *et al.,* 1991; Cooney *et al.,* 1991). Live recombinant adenovirus and poliovirus vaccine are all being under extensive studies (Prevec *et al.,* 1991; Cheng *et al.,* 1991; Minor *et al.,* 1990).

The V3 loop of the envelop protein gp120 is the principal neutralizing determinant (PND) of HIV, and an immunodominant cytotoxic T cell epitope has also been identified in this region (Takahashi *et al.,* 1988). However, it was recently shown that immunization with the PND was thought to induce only subtype specific immunity. However, it was recently shown that immunization with a 13-amino-acid residue peptide from this region induced antibodies that were cross-neutralizing from several HIV isolates (Javaherian *et al.,* 1990). Furthermore, a smaller hexapeptide (GPGRAF) from this region, when used to immunize rabbits, also led to antibodies capable of neutralizing divergent isolates. Analysis of the PND sequences of gp120 from 245 isolates of HIV-1 suggests that peptide cocktails of limited size may be feasible vaccine candidates. Of interest are recent attempts to construct a hybrid

synthetic peptide immunogen containing only those HIV epitopes necessary to induce a protective immune response, including antiviral CTL (Hart *et al.*, 1991). The complete hybrid consists of (from NH_2 to COOH terminus) the first 12-amino-acid residues of the gp41 fusion domain, a T cell epitope of gp120, a gp120 B cell epitope from the V3 loop region, and extending from this region, an additional five or six residue segment that comprises a CTL epitope. The resultant carrier-free hybrid peptide indeed had the ability to induce CD8+ MHC class I-restricted CTLs *in vivo* in mice. This finding demonstrate the potential of synthetic peptide-based vaccines to induce broad immune response toward HIV. However, because HIV can be transmitted as either a cell-free or cell-associated virus, a protective immune response against HIV will likely require both humoral and cell-mediated immunity.

Finally, in addition to the difficulties faced in developing an safe and effective anti-HIV vaccine, evaluating a candidate vaccine poses another part of difficulties. The first problems is the lack of indicators of protective immunity that can be followed. There presently are no strict correlations between the development or loss of certain types of immunity and the onset or progression of disease. The second obstacle is the prolonged period between a person's initial infection with HIV and the development of disease symptoms. This factor is particularly important in evaluating vaccines that use the killed whole virus or live attenuated virus as an immunogen, because the antibody profile (as measured by Western blot) that develop after immunization with such a vaccine cannot distinguish readily between a vaccine-induced seroconversion and a naturally acquired infection. The nature of the spread of HIV infection is a third obstacle to the evaluation of candidate vaccines. With an

infectious disease such as influenza, cohorts of volunteers can be immunized with a candidate vaccine in the autumn and, assuming an influenza outbreak in the winter, the efficacy of the vaccine can be established by spring. This situation is very different from the case of HIV infection in which transmission depends on behavioral factors and occurs over a long and variable period. Therefore, the time it would take to develop vaccine and then to prove its efficacy would be prolonged.

1.2 Evolution of HIV

Understanding the nature of sequence variation in the genome of human immunodeficiency virus is a crucial issue, because it is not merely relevant to unveiling the nature of viral pathogenesis and the human immune response but, more importantly, can provide information for the development of effective vaccines against HIV. Before the AIDS epidemic, our knowledge about retroviral variation was largely based on the studies of murine and avian oncoviruses. Recent comparisons of oncoviruses and the newly discovered retroviruses (HIVs, SIVs, and FIV) suggests that the immunodeficiency viruses and, more generally, the lentiviruses, may constitute a special case of viral and molecular evolution. High mutation rate with its consequent extensive diversity in their genomic sequences is probably the most significant feature of the lentivirus subfamily and that may relate to their highly replicative nature and well-recognized infidelity of reverse transcriptase (Preston *et al.*, 1988; Roberts *et al.*, 1988; Coffin, 1992). With the advent of DNA sequencing, and especially with the direct sequencing of PCR products without any *in vitro* manipulation (such as *in vitro* cloning), a large amount

of nucleotide sequence data has been obtained and errors generated by *Taq* polymerase during the amplification process can be readily avoided (Simmonds *et al.*, 1990*a*). By means of phylogenetic analysis, a lot of information has been extracted from these sequence data; information important in the understanding of the evolutionary processes of lentiviruses (Sharp *et al.*, 1988; Li *et al.*, 1988; Yokoyama *et al.*, 1988; Balfe *et al.*, 1990; Leigh Brown *et al.*, 1988; Holmes *et al.*, 1992).

In phylogenetic studies, the evolutionary relationships among viral strains, or among other groups of organisms are illustrated by means of a phylogenetic tree. Currently, there are a range of phylogenetic approaches available which can be used to construct evolutionary trees based on molecular sequence data. All these different phylogenetic approaches can be divided into two main categories: one is based on a statistical approach (UPGMA, Fitch-Margoliash, Neighbour-joining, Maximum Likelihood, etc.), and the other on the principle of parsimony (parsimony, invariants/evolutionary parsimony). Statistical approach category can further subdivided into Distance based methods (UPGMA, Fitch-Margoliash, Neighbour-joining, etc) and Maximum Likelihood method (Maximum Likelihood). Although different approach is based on different molecular evolutionary model (assumption about how the sequence have changed during evolution), they all try to solve the same problem: Having sequenced the 'same' DNA region (or 'same' protein) in several 'species', how do we then work out the phylogenetic relationships among the 'species'?

Assumptions related to different molecular evolutionary model are listed below (Wright *et al.*, 1993). From these assumptions, it is easy to see why

Maximum Likelihood model is currently the most favourite one and has been strongly recommended and widely used.

1) No variation in nucleotide frequencies during evolution: all methods assume this.

2) No between-lineages variation in rate of evolution: this assumption is used by the UPGMA method but most methods do not require this, e.g. Neighbour-joining, Fitch-Margoliash, Maximum Likelihood, parsimony and invariants/evolutionary-parsimony.

3) No between-site variation in rate of evolution: this assumption is make by most methods. Recent advances in Maximum Likelihood methodology allow for some rate variation along the sequence. In addition, one can construct a DNA distance measure that allows for different classes of site, with different rates for each.

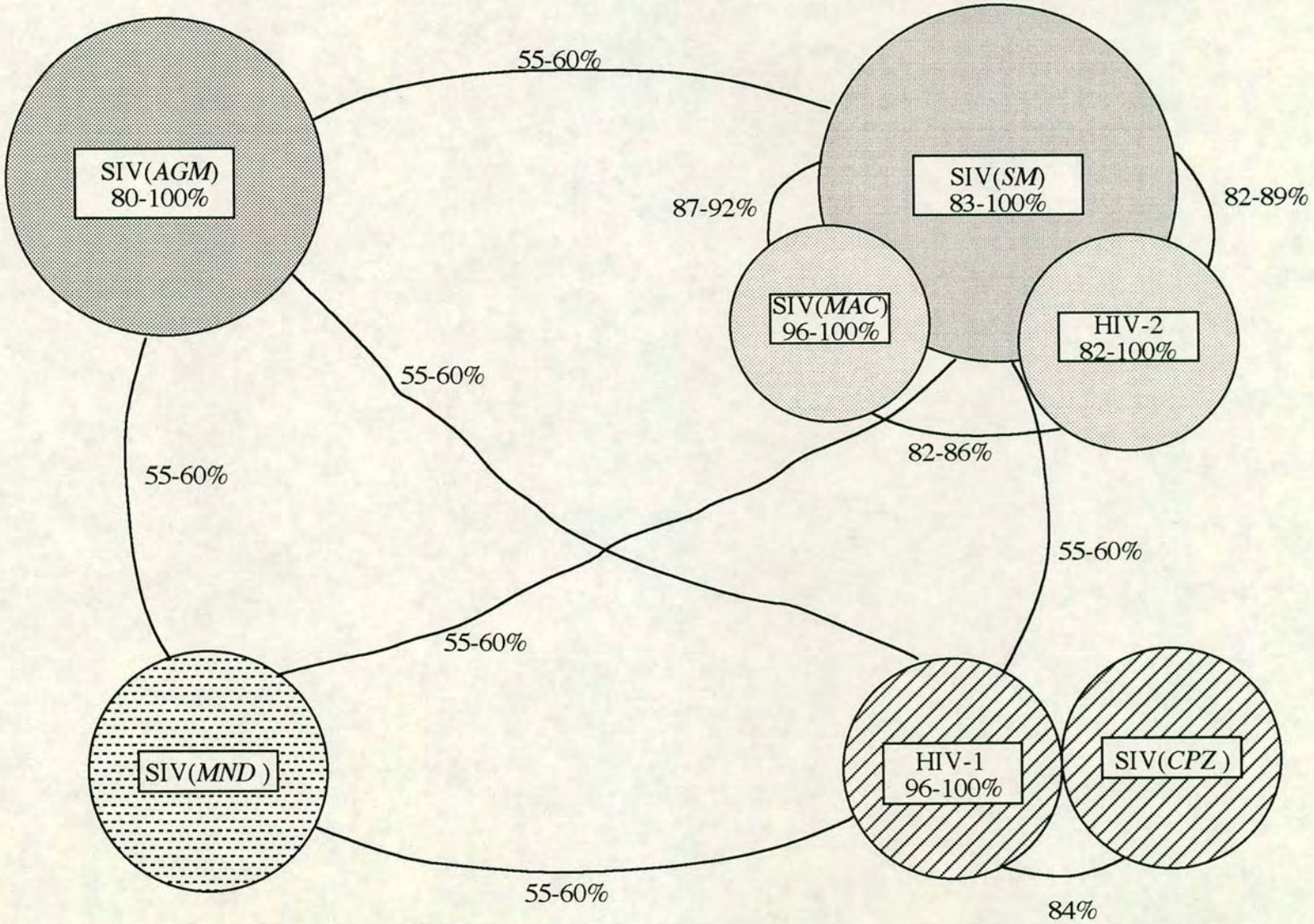
4) Sites are independent: this assumption is made by most methods. Recent advances in Maximum Likelihood methodology allow for 'patches' of correlated sites. No other methods at the moment (i.e. Feb. 1993) allow this.

As the likelihood principle is the predominant method of statistical inference used in this thesis, it is worthwhile spending a little time outlining the principle itself and how it has been used in the estimation of phylogeny. It should be emphasised at the outset that the statistical method of likelihood is not the critical element in the estimation of phylogeny; it is the particular model of molecular evolution that is the biological consideration and it is the model that my be reformed if we so desire. Likelihood estimation can proceed only on the basis of a given model. Likelihood requires three elements; a model (M), in this case the Felsenstein model of molecular

evolution, the data (D), the molecular sequences under comparison, and the competing hypotheses (H) which are simply the alternative tree patterns with associated time of divergence. There is always assumed to be a competing hypothesis in likelihood theory. Simply stated, likelihood theory states that hypotheses be judged on the basis of their likelihood. If P(D/H) is the probability of obtaining the data D given the hypothesis H, then the likelihood of an hypothesis H, given the experimentally determined data D, on the given model, is : $L_D(H) = P(D/H)$. To put it another way, likelihood assesses the probability of the data given the model under the various hypotheses - it does not assess the likelihood of the model. In addition, the relative merits of two competing hypotheses are contained in the likelihood ratio; $L_D(H1)/L_D(H2)$.

1.2.1 Evolutionary origin of HIVs and SIVs Lentiviruses have now been isolated and sequenced from a number of different species; sheep (Visna virus), goat (CAEV), horse (EIAV), cattle (BIV), cat (FIV), monkeys (SIVs) and humans (HIVs). Simian immunodeficiency viruses (SIVs) are non-human primate lentiviruses which to date have been isolated from macaques (SIV_{MAC}), sooty mangabeys (SIV_{SM}), African green monkeys (SIV_{AGM}), mandrills (SIV_{MND}) and a captive chimpanzee (SIV_{CPZ}) (Desrosiers, 1988 and 1990; Coffin, 1992; Feng *et al.*, 1992; Hirsch *et al.*, 1989*a, b* and 1990; Chakrabarti *et al.*, 1987). These five SIVs fall into four discrete groups based on genetic sequence analysis, with SIV_{MAC} and SIV_{SM} forming a single genetic group (Figure 1.4). SIVs are the closest known relatives of the HIVs, and each of HIV-1 and HIV-2 can be specifically associated with one of these different groups

Figure 1.4 Genetic grouping pattern of primate lentiviruses. Percentages on the lines are amino acid similarities in the *pol* gene, whereas those in squares are the degree of similarities within different isolates of each virus. (Desrosiers R.C. *Nature* (1990) Vol. 345, p288-289).


SIV(*AGM*)
80-100%
SIV(*SM*)
83-100%
SIV(*MAC*)
96-100%
HIV-2
82-100%
SIV(*MND*)
HIV-1
96-100%
SIV(*CPZ*)
55-60%
87-92%
82-89%
55-60%
55-60%
82-86%
55-60%
55-60%
55-60%
84%

of SIVs; SIV_{SM} in the case of HIV-2, and SIV_{CPZ} for HIV-1 (Desrosiers, 1990; Huet *et al.*, 1990; Hirsch *et al.*, 1989*a, b, c* and 1990; Feng *et al.*, 1992), thereby suggesting the possibility of the independent cross-species transfer of virus from simian hosts to humans. SIV_{AGM} isolates display much greater nucleotide sequence diversity than other SIVs and HIVs (Johnson *et al.*, 1989, 1990, 1991 and 1992; Allan *et al.*, 1991), and form a group as distinct as HIV-1/SIV_{CPZ} and HIV-2/SIV_{SM}. Furthermore, its high seroprevalence (up to 50%) in wild populations may suggest that SIV_{AGM} has been present in the African green monkey population for a long period of time (Johnson and Hirsch *et al.*, 1990, 1991 and 1992). SIV_{MND} is a close genetic relative of SIV_{AGM}, and therefore, the viruses related to the SIV_{MND} and SIV_{AGM} may serve as an ancestral virus of HIVs and SIVs (Desrosiers, 1990; Johnson *et al.*, 1992). The divergence time point between SIV_{AGM} and HIV, HIV-1 and HIV-2 has also been estimated to be between 40 to 280 years ago (Smith *et al.*, 1988; Li *et al.*, 1988; Sharp *et al.*, 1988). In general, SIVs from African non-human primates fail to cause AIDS-like disease in their natural hosts (Kraus *et al.*, 1989; Baier *et al.*, 1989; Johnson *et al.*, 1991 and 1992). However, cross-species transmission of viruses may result in disease. If the origin of the pathogenic SIV_{MAC} in Asian rhesus macaques was a recent cross-species transmission event from the sooty mangabey (Hirsch *et al.*, 1989*a*), and if the proposed genetic link between the HIV-1 and SIV_{CPZ} (Huet *et al.*, 1990), and HIV-2 and SIV_{SM} is firmly held (Hirsch *et al.*, 1989*c*), then pathogenesis would be closely associated with trans-species infection (Leigh Brown, 1991).

1.2.2 Extreme variability of Lentivirus Genome The lentiviruses, including HIV, appear to be among one of the most rapidly evolving genomic molecules. Lack of proof-reading exonuclease activity of their reverse transcriptase and a high rate of replication may jointly account for this extreme variability. Purified HIV-1 reverse transcriptase has recently been shown to be approximately ten-fold less accurate than avian reverse transcriptase (Roberts *et al.*, 1988; Preston *et al.*, 1988). Little is known about the biochemical properties of simian reverse transcriptase, however, considering its overall genomic mutation rate being similar to that of HIV (Johnson *et al.*, 1990 and 1992; Burns *et al.*, 1991; Baier *et al.*, 1990 and 1991; Desrosiers, 1990), the fidelity of simian reverse transcriptase could be similar. Genetic hypervariability has been suggested to contribute to the pathogenicity of lentiviruses simply because the host immune system can not cope with this extremely large number of virus variants (Nowak *et al.*, 1990 and 1991*a, b.*). However, recent analysis of the *in vivo* genetic variability of African green monkey SIVs has revealed that sequence diversity in these animals is considerably higher than that of HIVs (Johnson *et al.*, 1990 and 1992). This implies that SIV_{AGM} has been present in non-human primates much longer than in humans. As African green monkeys do not suffer severe pathological effects, hypervariability is unlikely to be the cause of viral pathogenicity (Leigh Brown, 1991).

Immune selection has been reported for many members of the lentiviral family (Clements *et al.*, 1980, 1982 and 1988). It is interesting to note that although the genetic organization of HIV's and SIV's is very similar (Baier *et al.*, 1990; Johnson *et al.*, 1992), there are differences in the location of the variable and conserved

regions of gp120 (Johnson *et al.*, 1992). This is of particular importance if such changes also involve the immunogenic epitopes. This raises a number of possibilities; 1) whether difference in epitope location between HIVs and SIVs is a consequence of cross-species transmission; 2) is there any relationship between this difference and pathogenesis in non-natural hosts and, 3) is the immune response in non-human primates so different from that of humans that SIV infected monkeys should be reconsidered as viable animal models.

1.2.3 Sequence Diversity in the *env* Gene of HIVs As mentioned above, HIV shows considerable sequence variation. Such sequence diversity is not only observed between independent isolates (Hahn *et al.*, 1986; Meyerhans *et al.*, 1989; Alizon *et al.*, 1986; Starcich *et al.*, 1986), but also among sequential isolates from the same individual patient (Hahn *et al.*, 1986; Simmonds *et al.*, 1990*a* and 1991). A comparison of published sequences has suggested that the rate of substitution is not constant across the entire HIV genome, but is particularly high in the hypervariable regions of the *env* gene (14×10^{-3} nucleotide substitutions per site per year), and lower in the *gag* and *pol* regions (around 1.7×10^{-3} nucleotide substitutions per site per year) (Li *et al.*, 1988). These results suggest that the HIV *env* gene, in general, is evolving under different selective constraints from the rest of the viral genome. Within the *env* gene, there is also a distinct pattern of variation. Five hypervariable regions of gp120 (V1-V5) were defined as regions with less than 25% conservation of amino acids between several published sequences (Modrow *et al.*, 1987).

Recently, a principal neutralization determinant (PND) of HIV-1 has been identified

to localize in a loop structure within the third hypervariable region (V3) (Rusche *et al.*, 1988; Palker *et al.*, 1988; Javaherian *et al.*, 1989), suggesting that immune selection may play an important role in the generation of sequence variability in the V3 region. This notion was supported by recent experiments showing an *in vivo* emergence of virus variants that were resistant to neutralization by autologous sera (Albert *et al.*, 1990; Montefiori *et al.*, 1991). The immunogeneity and functions of the other hypervariable regions have not yet been studied as extensively as those of the V3 region.

Located between the hypervariable regions are sequences which are highly conserved or with intermediate variability. Some parts of these conserved sequences have been shown to be important for viral function. For example, two conserved regions located at the amino (residues 36 to 45) and carboxyl (residues 491 to 501) termini of the gp120 glycoprotein are necessary for the noncovalent association with gp41 transmembrane glycoprotein (Helseth *et al.*, 1991). Of course, the major CD4 binding domain located between the V4 and V5 hypervariable regions is the most studied. Monoclonal antibodies to the CD4 receptor and soluble CD4 molecules are both effective in blocking virus infection, suggesting that binding to this portion of the CD4 molecule through viral exterior glycoprotein is a vital step in the infection process (Putney *et al.*, 1990*a, b*). Conserved sequences within the hypervariable regions also have indispensable functions. This is especially true for the amino acids Gly-Pro-Gly which are located at the tip of the V3 loop bonded by two cysteins. A single change in the amino acid sequence of these sequences can generate a virus that has either changed its tropism characteristics or substantially reduced its infectivity

(Cordonnier *et al.*, 1989; Bolognesi, 1990; Takeuchi *et al.*, 1991). In addition, twenty-one cysteine residues in the envelope gene are also completely conserved in all isolated reported which suggests the vital role of these cysteine residues for envelope functions (Tschachler *et al.*, 1990; Dedera *et al.*, 1992). The relatively conserved region in the transmembrane glycoprotein gp41 is also believed to play a role in virus infection. This region, defined as the fusogenic domain, is thought to reside within the N-terminal portion of gp41 (Gallaher, 1987; Gonzalez-Scarano *et al.*, 1987). Fusion occurs after virus binding to the CD4 molecule. This process can also occur between virus-infected cells exhibiting gp120 and gp41 on their surface and uninfected cells bearing CD4 molecules (Lifson *et al.*, 1986*a, b*; Sodroski *et al.*, 1986*a*). This results in the formation of multinucleated giant cells and represents a form of virus cytopathogenicity. At the same time, this process also allows HIV to pass directly from cell to cell.

Whereas hypervariability in the *env* gene has been observed in samples from within individuals, and especially from those in the later stages of infection, the *in vitro* propagation of viral isolates however, significantly reduces the genetic diversity in the V3, V4 and V5 regions (Meyerhans *et al.*, 1989; Kusumi *et al.*, 1992). It has been shown that genetic diversity can drop to zero after *in vitro* culturing for about 10 weeks (Simmonds, personal communication), suggesting some sort of selection on the envelope gene. In addition, such homogeneity of isolated viruses also confirms the role that immune selection plays in the generation of virus variants *in vivo*.

1.2.4 Sequence Diversity In the *gag* and *pol* Genes of HIVs Studies of sequence variation have been concentrated on the recognised hypervariable regions of the *env* gene, particularly on the V3 region simply because it is highly antigenic and is indispensable for virus functions (see section 1.2.3). It was not until recently that attention has been drawn to the *gag* and *pol* genes. Within the *gag* gene the pattern of sequence variation is not random. The p15 and p17 *gag* proteins appear to evolve quite quickly at amino acid replacement sites, whereas the major core protein, *gag* p24 remains more conserved (Hahn *et al.*, 1986; Coffin, 1992). The degree of sequence variation seen in some parts of the *gag* p17 or p15, may be very useful to track down the epidemiological and transmissive information of HIV because these two regions can provide enough variation information to distinguish different possible source of infection and, at the same time are not under sever selection which may cause convergent evolution (Holmes, personal communication). Phylogenetic analysis based on the *gag* nucleotide sequences has revealed a grouping pattern between different members of subfamily lentivirinae, consistent with their geographic and species collocation (Li *et al.*, 1988; Myers *et al.*, 1991). Moreover, recent results obtained in our laboratory suggested that phylogenetic analysis based on the *gag* p17 nucleotide sequences can determine contact network between individuals within an infected community (Dr. E. Holmes, personal communication). Thus, an understanding of sequence variation can also help us to further understand the transmission and epidemiology of HIV-1.

Several immunogenic epitopes have been identified in both the *Gag* and *Pol* proteins which can be recognised either by neutralizing antibodies or by cytotoxic T

cells (Hosmalin *et al.*, 1989; Papsidero *et al.*, 1989; Walker *et al.*, 1989). Phillips *et al.*, demonstrated that sequence variation in one of the *gag*17 CTL (cytotoxic T lymphocytes) epitopes (residue 21-35, L R P G G K K K Y K L K H I V) can lead to loss of CTL recognition, and suggested this could be one way by which HIV escapes immune surveillance (Phillips *et al.*, 1991). In general, different HLA (Human Leucocyte Antigens) class I molecules select distinct epitopes derived from HIV proteins to stimulate CTL responses. Therefore, apart from the HIV sequence itself, HLA type could also have an impact on virus escape and possibly clinical progress. The relatively high frequency of HIV amino acid replacement observed in HLA B8 patients, compared with HLA B27 patients, suggests that the HLA B8-restricted *Gag* epitope(s) are less constrained and that amino acid changes in this regions may result in CTL-escape (Phillips *et al.*, 1991). To some extent, this result can be used to explain the association between HLA haplotype (and especially A1 B8 DR3) and the outcome in HIV infection seen in Edinburgh haemophiliac patients. Relative faster progression towards AIDS among the members of the Edinburgh haemophiliacs is probably associated with their A1 B8 DR3 haplotype (Steel *et al.*, 1988). A1 B8 DR3 is by far the commonest haplotype in caucasians, with a particularly high frequency among those of northern European descent (Steel *et al.*, 1988).

Immune recognition and immune escape mutants have not been studied as extensively in *pol* as that in *env* and *gag* genes. However, studies of AZT (3'-azido-3'-deoxythymidine), ddI (2', 3'-dideoxyinosine) and ddC (2', 3'-dideoxycytidine) resistant virus variants highlight the sequence variation in the *pol* gene. Larder *et al.*,

firstly reported that isolates of HIV taken from patients who had been taking AZT for more than 6 months, were resistant to AZT *in vitro* (Larder *et al.*, 1989*a, b, c*). Analysis of the coding region of reverse transcriptase (RT) from clinical isolates by nucleotide sequencing led to the discovery that multiple common nucleotide changes were associated with this resistance. These changes conferred specific amino acid replacements in RT at the following residues: Asp67 -> Asn, Lys70 -> Arg, Thr215-> Phe or Tyr, and Lys219 -> Gln (Larder *et al.*, 1989*a, b, c* and 1990). The most resistant isolates had mutations in all four residues or the first three (Richman *et al.*, 1992). Recently, one more substitution at position 41 (Met -> Lys) in the RT region was found to contribute to AZT resistance (Kellam *et al.*, 1992). Moreover, mutation in the RT regions has also been shown to be responsible for ddI resistant virus strains. The residue replacement at position 74 is one of the putative sites for this resistant effect (St. Clair *et al.*, 1991). However, it is still not clear whether the resistant strains are as pathogenic and virulent as the wild strains, how the proportion of the resistant strains correlates with the disease stage and what determines the rate of development of resistant strains. Finally, in this work (see Chapter 3), the observation of one substitution at position 70 (Lys -> Arg) in sequence data generated from one of the factor VIII concentrates prepared before the use of AZT as an antiviral agent suggests that there was a pre-existing polymorphism (Zhang *et al.*, 1991).

1.3 Previous Studies of Sequence Evolution in the Edinburgh Haemophiliac Cohort

Our research group have been studying sequence variation in a group of haemophiliac patients who became infected between March and May in 1984 after exposure to a single common batch of HIV-contaminated factor VIII concentrates, prepared from locally donated blood by the Scottish National Blood Transfusion Service (SNBTS) (Ludlam *et al.*, 1985). Of a total of 32 patients exposed to the batch, 18 became HIV seropositive during the subsequent ten months. Their seropositivity was tested by a range of enzyme-linked immunosorbent assays (ELISAs) detecting antibodies to different components of the virus and confirmed on western blotting (Simmonds *et al.*, 1988). These seropositive patients received, on average, significantly more units of the putative infected batch than those who remain seronegative (Ludlam *et al.*, 1985).

Samples donated from eight patients (patients 74, 77, 79, 82, 83, 84, 87 and 91) from this cohort, and from one non-cohort HIV-infected haemophiliac (patient 12) were studied first (Simmonds *et al.*, 1988 and 1990*a*). A double polymerase chain reaction procedure based on nested primers was developed in our laboratory and has been used to detect and quantify the provirus DNA in the peripheral blood mononuclear cells (PBMCs) (Simmonds *et al.*, 1990*a*). The PCR amplified products were then Gene-cleaned to remove the non-incorporated nucleotide triphosphates and primers. Direct sequencing method was subsequently applied to the purified PCR products without *in vitro* cloning procedures. Direct sequencing approach can avoid obtaining bias sequences from the *in vitro* virus culture and artificial mutated

sequences which is due to errors introduced by *Taq* polymerase during the amplification process (Balfe *et al.*, 1990; Simmonds *et al.*, 1990*a* and 1991).

Phylogenetic analysis of nucleotide sequences from *gag* p24, and the V4-V5 and V3 regions of *env* gene revealed similar phylogenetic patterns; 1) sequences from the same patient are more similar to each other than to those from different patients, 2) the non-cohort haemophiliac patient, who was infected in the United States from commercial factor VIII, has sequences distinct from those of the cohort patients, as are the published sequences from HIV viral isolates, 3) in the V4-V5 region, six of the eight Edinburgh-infected haemophiliacs (patients 77, 79, 83, 84, 87, and 91) have sequences which are particularly closely related to each other, whereas two others (patient 82, and 74) have sequences which are less so (Balfe *et al.*, 1990; Simmonds *et al.*, 1990*a*). According to the most recent sequence analysis based on the p17 region of the *gag* gene, it is likely that more than one virus variant was present in the original batch of factor VIII concentrates (Dr. E. Holmes, personal communication).

1.4 The Outline Research Presented In This Thesis

The research outlined in this thesis was primarily designed to study the temporal changes of virus load and of viral RNA sequences during the course of infection. Accordingly, a highly sensitive and reliable RNA PCR method was first developed to enable the detection and quantification of viruses directly from plasma and serum of HIV-1 positive individuals and from blood products such as factor VIII concentrates (Zhang *et al.*, 1991). With the advent of direct sequencing of PCR

products, viral RNA sequences, in particular the V3 and V4 regions of the envelope gene, have been readily obtained from serial plasma samples taken from a single HIV-1 infected haemophiliac, and from the plasma samples of 5 independently infected patients prior to or immediately after seroconversion. By analyzing the temporal spectrum of quantitative and qualitative features of plasma viral RNA sequences, it has been possible to assess the contribution of viral load to the pathogenesis of HIV-1 and patients' clinical outcome; to improve our understanding of the process and pattern of viral genetic evolution; and by comparing with those sequences obtained from proviral DNA sequences present in PBMCs, to trace the viral genetic flow between these two distinct populations and compartments. Furthermore, by revealing the genetic features of plasma viraemia at early stage of the infection, the nature of viral replication before the immune response is mounted can be better understood which, in turn, is helpful to assess the effects and the consequences of the human immune response.

My research project, and indeed this thesis, is generally separated into three parts. The first part (Chapter 3) is exclusively devoted to the development of a highly sensitive and reliable RNA PCR technique which can be used to detect, quantify and sequence cell-free HIV RNA directly from the plasma or serum of HIV-1 positive patients (Zhang *et al.*, 1991). Using this RNA PCR based technique, plasma from 10 out of 12 haemophiliac patients tested were found to contain detectable levels of HIV-1 RNA. On average, HIV RNA was more abundant in the plasma of patients with more advanced disease compared to asymptomatic patients. However, in accordance with others findings (Ho *et al.*, 1989; Coombs *et al.*, 1989;

Schnittman *et al.*, 1991), the most striking results obtained from the viral quantification study of these haemophiliacs is the high level of cell-free HIV RNA found in some of the CDC group II patients (Zhang *et al.*, 1991), indicating the persistence of viral replication throughout the whole course of an HIV infection and, therefore, suggesting that there is no virological 'latent' period to correspond with that observed in the clinical outcome. At the same time, this finding may also suggest that early treatment of the HIV-1 infected patients is desirable.

The second part of the work (Chapter 4) is an extensive study of sequential sequence variation in the HIV envelope gene from a single HIV-1-positive haemophiliac (patient 82). Starting from seroconversion (1984), six serial plasma samples were collected at year 3 (1987), 4 (1988), 5 (1989), 6 (1990) and 7 (1991). Sequence analysis was largely concentrated on the V3 and V4 hypervariable regions of the envelope gene. A total of 89 V3 and 114 V4 viral RNA sequences were obtained directly from plasma samples without any *in vitro* manipulations. Phylogenetic analysis of these sequences reveals that there are complex evolution patterns in these two hypervariable regions. For the V3 regions, all subsequent HIV genotypes can be explained on the basis of the accumulation of sequence differences from the genotype found at seroconversion (Holmes *et al.*, 1992). A major diversification of V3 genotypes had taken place in the plasma within 3 years from seroconversion. Several distinct lineages could be identified at year 3, but only two persisted and subsequently reached high frequency. Furthermore, selective constraints on the V3 regions, particularly in the V3 loop region, were clearly observed which was indicated by the dramatic fluctuations in the frequency of sequence variants and

the extensive convergent evolution (identical amino acid changes occurring in independent lineages) (Holmes *et al.*, 1992). These findings strongly suggest that although there is selection for replacement of amino acids which may alter the B and T cell epitopes once they are recognised, there is also a severe selective constraint as to which amino acids are functionally viable within these regions. In addition, because of the extensive constraints and convergent evolution of the V3 sequences, any interpretation of epidemiological relatedness based on this region may not be appropriate.

Sequence variation in the V4 region is slightly different. Apart from frequent amino acid replacement through time, length variation is also observed in this region. Rapid sequence change, consisting of regular replacement by a succession of distinct viral populations, was found in the plasma (Simmonds *et al.*, 1991). Each succeeding sequence type is not obviously more related to those that come before or after it than they are to the sequences of the original infecting viruses, suggesting the evolution of viral RNA V4 sequences in the plasma population of this patient is discontinuous and may be largely contributed by the hidden evolution in the solid tissues such as lymph node, brain, spleen, lung, liver and etc. More interestingly, there appears to be significant differences between the frequencies of sequence variants in DNA and RNA populations within the same sample, indicating that at any one time point, the predominant plasma virus variants were antigenically distinct from those viruses encoded by HIV DNA sequences in PBMCs. How these findings contribute to our understanding of HIV infection and pathogenesis is discussed in Chapter 4.

The transient high level of viraemia present in patient plasma around

seroconversion has been known for some time (Goudsmit *et al.*, 1986; Gains *et al.*, 1987; McRae *et al.*, 1991), however, it was not until 1991 that the quantitative assay of viruses present in plasma population at this stage of infection was published (Daar *et al.*, 1991; Clark *et al.*, 1991). In the final part of my work (Chapter 5), attention is drawn exclusively to the study of sequence variability of the envelope gene at the first stage of infection. Plasma samples were collected from 5 independently infected patients prior to or immediately after seroconversion. Two of these five patients were infected through sexual contact (Sc1 and Sc2) and the other three infected parenterally (p82, p74 and p84). PBMC samples were collected 3-6 months after seroconversion from 4 members (p28, p79, p77 and p84) of the Edinburgh haemophilic cohort. The sequence analysis of V3, V4 and part of p17 region of the *gag* gene reveals very different pictures of sequence variability. Contrary to what is observed in later stages of infection where considerable sequence variation in envelope gene is readily demonstrated (Balfe *et al.*, 1990; Simmonds *et al.*, 1990*a* and 1990; Holmes *et al.*, 1992), no sequence variation was found in the V3 region in any of the samples. The closely linked V4 region was also lacking in variation in all but one (p82) cases investigated. Unexpectedly, sequence variation *is* present in the *gag* gene at this stage. Furthermore, 3 out of 4 members of the Edinburgh haemophiliac cohort have identical V3 and V4 proviral sequences and the fourth patient has only one nucleotide difference in the V3 region from the others. This striking finding suggest that V3 region sequences, and possibly part of V4 sequences, are under selection at transmission of HIV. As the V3 region contains one of the major targets for both B and T cells, identifying those sequences which are selected

for in the early stage of viral infection will be important not only in terms of documenting the viral sequence changes which can overcome the specific immune response, tracing those mutations at the molecular level which may determine the changes in viral tropism, but also in providing sequence information for the design and development of an effective vaccine.

CHAPTER 2

MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Blood Samples

2.1.2 Plasma or Serum and PBMCs Samples

2.1.2.1 For Viral RNA Quantification Study

2.1.2.2 For Sequential Sequence Variation Analysis

2.1.2.3 For Sequence Variation Study During the Primary Stage of HIV-1 Infection

2.1.3 Factor VIII Concentrates

2.1.4 HIV and tk Primers

2.1.5 pBH10.R3

2.1.6 pSV2gpt

2.1.7 Virus Isolates

2.1.8 Cell Line Used in *in vitro* Virus Culture

2.2 METHODS

2.2.1 Separation and Storage of Plasma and PBMC from Blood

2.2.2 *In Vitro* Culturing HIV_{HXB2}, $HIV_{HTLV\text{-}IIIB}$, HIV_{RF} Variants and Harvesting Culture Supernatant

2.2.3 DNA Extraction from PBMCs

2.2.4 Viral RNA Extraction from Plasma, Factor VIII Concentrates and Cell Culture Supernatant

2.2.5 Reverse Transcription of Viral RNA

2.2.6 Measurement of Reverse Transcriptase Reaction Efficiency by

2.1 MATERIALS

2.1.1 Blood Samples Seropositive blood samples from HIV-1 infected haemophiliacs, intravenous drug users and seronegative individuals with no known risk-factors for HIV infection, were obtained from Drs. Christopher Ludlam (Edinburgh Royal Infirmary) and Roy Robertson (Edinburgh City Hospital). Plasma and peripheral blood mononuclear cells (PBMCs) were separated by centrifugation over Ficoll lympaque (Nycomed). Plasma samples were stored at -70^{0}C while PBMCs were in liquid nitrogen (see section 2.2.1).

2.1.2 Plasma or Serum and PBMCs Samples

2.1.2.1 For Viral RNA Quantification Study Plasma samples were obtained from 12 HIV-1 infected haemophiliacs who seroconverted in 1984 following transfusion by a single common batch of HIV-contaminated factor VIII (Ludlam *et al.*, 1985). All 12 individuals seroconverted for antibody between 3-10 months after receiving the factor VIII (Simmonds *et al.*, 1988). Patients 56, 70, 82, 83 and 84 have been classified as CDC (Centre for Disease Control) stage II. Patients 72, 74, 77, 79, 87 and 95 are in CDC stage IV at the time of study (in year 1989). These CDC stage IV patients have been suffering from a range of opportunistic infections and constitutional symptoms of HIV infection. Patient 28 died in 1988. Apart from patient 72, all CDC IV patients but none of the CDC II patients have been receiving antiviral treatment (AZT) (Dr. Henry Watson, Department of Haematology,

Edinburgh Royal Infirmary, personal communication).

2.1.2.2 For Sequential Sequence Variation Analysis Six plasma samples (March, 1984; June, 1987; January, 1988; February, 1989; April, 1990 and January, 1991) from a single haemophiliac patient (patient 82) were collected and stored at -70^{0}C before viral RNA extraction. Patient 82 was asymptomatic at the time of study and has never undergone any antiviral therapy but has persistently low CD4 counts of less than 0.2 x 10^9 per litre blood (Dr. Henry Watson, Department of Haematology, Edinburgh Royal Infirmary, personal communication).

2.1.2.3 For Sequence Variation Study During the Primary Stage of HIV-1 Infection Seroconversion plasma samples from 5 independently infected patients (p82, p74, p84, Sc1 and Sc2) and three follow-up plasma samples (p82, p74 and p84) were obtained from Drs. Roy Robertson (Edinburgh City Hospital), Christopher Ludlam (Edinburgh Royal Infirmary), G.E.D. Urquhart, and A.J. France (Department of Medical Microbiology, University of Dundee). Patient 82, 74 and 84 seroconverted between March and May, 1984 whereas patient Sc1 and Sc2 in 1990 and 1991, respectively. No patient except patient 74 is currently on, or previously received any anti-viral treatment. Two (Sc1 and Sc2) of these five patients were infected through sexual contact and the remaining three parenterally (p82, p74 and p84). PBMC samples were collected 3-6 months after seroconversion from 4 members (p28, p79, p77 and p84) of the Edinburgh haemophilic cohort who were infected from a single common batch of HIV-1 contaminated factor VIII concentrates

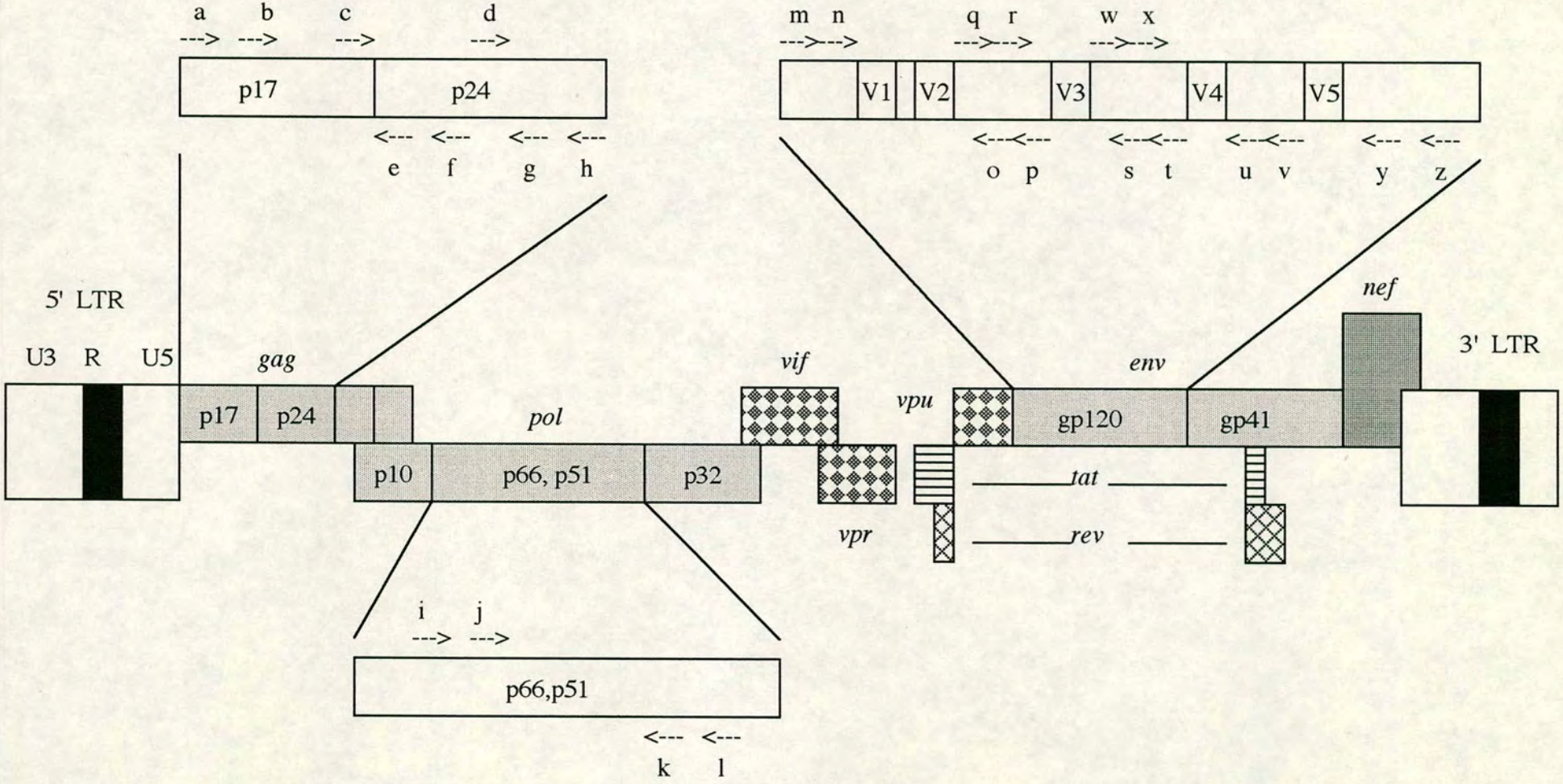
in 1984 (Ludlam *et al.*, 1985). One of the PBMC samples was collected from the same patient (p84) from whom the plasma sample was also obtained.

2.1.3 Factor VIII Concentrates Eight batches of factor VIII concentrates, including both commercial and NHS (National Health Service)-produced material, were obtained from Dr. G. Kemball-Cook at the National Institute of Biological Standards (NIBSC). All batches were unheated and prepared before the introduction of donor screening for anti-HIV antibodies. They were distributed in the United Kingdom between 1981 and 1984. All factor VIII concentrates were reconstituted according to the manufacturer's instructions.

2.1.4 HIV and tk Primers HIV primers were synthesised by the Oswel DNA Service, Department of Chemistry, University of Edinburgh, and were purified by HPLC. The primers were based on the consensus of several published HIV sequences (HIV_{JB}, HIV_{HXB2}, HIV_{ELI}, HIV_{LAI}, HIV_{MAL}, HIV_{MN} and HIV_{RF})[1]. The primer sites were chosen for greatest conservation between these published sequences. No more than one mismatch with any of the above published HIV sequences was permitted and nor was any mismatch near the 3' terminus. The positions of these HIV primers along the HIV_{HXB2} genome are schematically presented in Figure 2.1. The primers for HSV (herpes simplex virus) tk gene and HCV (hepatitis C virus) *NS3*

[1]According to the newly revised naming system in the HUMAN RETROVIRUSES and AIDS 1991 database by Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

Figure 2.1 Schematic illustration of HIV primers' positions along the HIV_{HXB2} genome. Those above indicated regions such as a, b, c, d,..., are sense primers whereas those underneath such as e, f, g, h, ..., are anti-sense ones.


a
b
c
d
p17
p24
e
f
g
h
m
n
q
r
w
x
V1
V2
V3
V4
V5
o
p
s
t
u
v
y
z
5' LTR
U3
R
U5
gag
p17
p24
p10
pol
p66, p51
p32
vif
vpr
vpu
env
gp120
gp41
nef
3' LTR
tat
rev
i
j
p66,p51
k
l

region were kindly provided by Dr. R. Al-Shawi (Division of Biological Sciences, University of Edinburgh) and Dr. Peter Simmonds (Department of Medical Microbiology, University of Edinburgh), respectively (Al-Shawi *et al.*, 1988; Simmonds *et al.*, 1990). The sequences of the primers for HIV template, the HCV NS3 region and transcribed HSV tk template are given below and the coordinates listed in the brackets of HIV primers are from the HIV_{HXB2} sequence (+ : sense; - : anti-sense).

HIV *gag* primers:

(a) 5'GCGAGAGCGTCAGTATTAAGCGG, (+, 795),

(b) 5'GGGAAAAAATTCGGTTAAGGCC, (+, 835),

(c) 5'GGTACATCAGGCCATATCACC, (+, 1214),

(d) 5'GAGGAAGCTGCAGAATGGG, (+, 1407),

(e) 5'CTTCTACTACTTTTACCCATGC, (-, 1248),

(f) 5'TCTGATAATGCTGAAAACATGGG, (-, 1296),

(g) 5'GGTCCTTGTCTTATGTCCA, (-, 1636),

(h) 5'ACCGGTCTACATAGTCTC, (-, 1669).

HIV *pol* primers:

(i) 5'CCCAAAAGTTAAACAATGGCC, (+, 2602),

(j) 5'AGAAATTTGTACAGAGATGG, (+, 2653),

(k) 5'CCATTTATCAGGATGGAGTTC, (-, 3245),

(l) 5'GCTGTCTTTTTCTGGCAGCAC, (-, 3281).

HIV *env* V1-V2 primers:

(m) 5'GAGGATATAATCAGTTTATGG, (+, 6577),

(n) 5'GATCAAAGCCTAAAGCCATG, (+, 6599),

(o) 5'CAATAATGTATGGGAATTGG, (-, 6930),

(p) 5'GTACATTGTACTGTGCTGACA, (-, 7020).

V3-V4 primers:

(q) 5'TACAATGTACACATGGAATT, (+, 6957),

(r) 5'TGGCAGTCTAGCAGAAGAAG, (+, 7009),

(s) 5'CTGGGTCCCCTCCTGAGG, (-, 7314),

(t) 5'ATTACAGTAGAAAAATTCCCC, (-, 7361),

(u) 5'ATTCTGCATGGGAGTGTG, (-, 7465),

(v) 5'GGAGGGGCATACATTGC, (-, 7520).

V4-V5 primers:

(w) 5'TCAGGAGGGGACCCCAGAAATT, (+, 7316),

(x) 5'GGGGAATTTTTCTACTGTAAT, (+, 7361),

(y) 5'CTTCTCCAATTGTCCCTCATA, (-, 7645),

(z) 5'GCCCATAGTGCTTCCTGCTGCT, (-, 7795).

HCV *NS3* primers:

(EDI) 5'GTGGTCGACTGCAATACGTGTGTCAC (+),

(ED2) 5'CCGGCATGCATGTCATGATGTAT (-),

(ED3) 5'CACCCAGACAGTCGATTTCAG (+),

(ED4) 5'GTATTTGGTGACTGGGTGCGTC (-).

HSV tk primers:

(1) 5'GCCAGTAAGTCATCGGCTCGGG (+),

(2) 5'CCATCAACACGCGTCTGCGTTCG (-).

2.1.5 pBH10.R3 is a recombinant plasmid containing a nearly full length HIV-1 genome (Simmonds *et al.*, 1990*a*). Cesium chloride (CsCl) gradients-purified plasmid was quantified by absorbance measurement at 260 nm.

2.1.6 pSV2gpt is also a recombinant plasmid containing a mouse promoter region and coding region of the herpes simplex virus type 1 thymidine kinase gene (HSV tk) (Al-Shawi *et al.*, 1988).

2.1.7 Virus Isolates

HIV_{HXB2} is an infectious molecular clone obtained from a phage library of HIV-1 DNA obtained from HIV-1 infected H9 cells (Shaw *et al.*, 1984).

$HIV_{HTLV\text{-}IIIB}$ / HIV_{LAI} isolate originated from a French AIDS patient LAI. This isolate used to be called LAV-1 to distinguish it from HIV2 (LAV-2) and formerly designated HIV_{BRU} (Wain-Hobson *et al.*, 1991).

HIV_{RF} isolate also designated as HAT because the virus was isolated from a patient of Haitian descent (Myers *et al.*, 1991).

2.1.8 Cell Line Used in *in vitro* Virus Culture

C8166 cells are a T cell line containing a genome of HTLV-I that expresses only the *tat* gene (Sodroski *et al.*, 1984).

2.2 METHODS

2.2.1 Separation and Storage of Plasma and PBMC from Blood About 20mls of heparinized whole blood sample was received at each time point and the separation of plasma and PBMC normally followed within 2 hours after arrival. Two 10mls of whole blood were layered over two 10mls of Ficoll lymphopaque (Nycomed) and then spun at 2,500 rpm for 15 minutes with no brake. The plasma from the top phase was aliquoted into Nunc freezing cryo vials in 1ml amounts and frozen at -70^{0}C. The interphase of PBMC was aspirated off and washed twice with 20ml RPMI 1640 (Gibco). Approximately 5 x 10^6 cells were resuspended in 1 ml of freezing media (1 volume RPMI 1640 supplemented with 10% fetal calf serum (FCS), 1% penicillin, streptomycin and 1 volume 20% DMSO in 80% FCS) and the mixture was put into a Nunc tube and stored at -70^{0}C. Two days later, the vials were placed in the appropriate section in the liquid nitrogen store (vapour phase).

2.2.2 *In Vitro* Culturing HIV_{HXB2}, $HIV_{HTLV-IIIB}$, HIV_{RF} Variants and Harvesting Culture Supernatant The C8166 cells infected with HIV variants (HXB2, HTLV-IIIB and RF) were cultured at 37°C in PRMI 1640 (Gibco) medium supplemented with 20mM L-glutamine, 10% heat-inactivated FCS, 50 units/ml penicillin and 50µg/ml streptomycin. The culture flasks were checked under the inverse microscope (Nikon) for the presence or absence of syncytia, and the cultured supernatant in which syncytia was formed was collected and stored in -70°C for subsequent viral RNA extraction and quantification. All HIV culturing work was carried out in the Category 3 laboratory.

2.2.3 DNA Extraction from PBMCs There are a number of procedures which have been developed to extract DNA from cells and clinical specimens. The following procedure which has been used in this work represents a reasonable compromise between quality of DNA and ease of use, yielding fairly pure high molecular weight DNA, which can be stored at -20°C without obvious degradation (Simmonds *et al.*, 1990). Pelleted cells (approximately 5 x 10^6) were suspended in 400µl of lysis buffer (50mM Tris-HCl pH8.0; 50mM EDTA pH8.0; 100mM NaCl; 0.01% w/v Proteinase K; 1% N-lauroylsarcosine) and incubated at 65°C for 2 hours. Phenol and chloroform extraction followed and DNA was precipitated with two volumes of 100% ethanol at -20°C for 2 hours. The precipitated DNA was collected by centrifugation and then dried at 50°C for 10-15 minutes. Finally the DNA was resuspended in 20-200µl of distilled water and DNA concentration and purity were assessed by spectrophotometry at an absorbance of 260 and 280nm.

2.2.4 Viral RNA Extraction from Plasma, Factor VIII Concentrates and Cell Culture Supernatant Considerably more care is required to successfully extract RNA from plasma samples, factor VIII concentrates and cell cultured supernatant in view of its greater sensitivity to degradation than DNA. 500µl of patient plasma or 500µl of HIV infected C8166 culture supernatant was mixed with 8.5ml of phosphate buffered saline (PBS, pH7.3); alternatively, 3ml of factor VIII reconstituted with the recommended volume of water was mixed with 6ml of PBS; in both cases, virus was pelleted at 45,000 x g in a swing out rotor (Sorvall SH80) at 4^{0}C for 2 hours (centrifuge Sorvall RC28S). The pellet was resuspended in 1.2ml of a denaturing solution (2M guanidinium thiocyanate; 12.5mM sodium citrate pH7.0; 0.25% sarcosyl; 0.05M 2-mercaptoethanol, 50% water-saturated distilled phenol and 1ug carrier RNA) and mixed thoroughly with 200µl chloroform (Chomczynski *et al.*, 1987). After vigorous shaking for 15 seconds, the solution was incubated on ice for 15 minutes. The sample was then spun at 14,000 x g for 15 minutes and the aqueous phase, which contained the HIV RNA, was precipitated with an equal volume of isopropanol for at least 45 minutes at -20^{0}C. Precipitated RNA was pelleted by spinning for 15 minutes at 14,000 x g at 4^{0}C. The pellet was washed once with 1ml 75% ethanol, dried under vacuum for 10 minutes and dissolved in 20µl of autoclaved RNase-free Analar water (BDH). Sometimes, a second phenol-chloroform extraction was necessary to remove remaining protein, and this was carried out either after aqueous phase transfer or after dissolution of the RNA pellet. If HIV was present in a large amounts (in long term culture supernatant for instance) the high speed centrifugation step could be eliminated simply by mixing 250µl supernatant with

750µl denaturing solution and 200µl chloroform in the first step, and using the same procedures as before in subsequent steps. The presence of carrier RNA, which may protect HIV RNA from RNase degradation and stabilize reverse transcriptase was critical in the subsequent DNase digestion and cDNA synthesis steps. The source of carrier RNA was probably not critical, total cellular RNA from either adult mouse liver or cultured sheep fibroblast cell lines have been used with equivalent results.

2.2.5 Reverse Transcription of Viral RNA The successful PCR (polymerase chain reaction) amplification of viral RNA is largely dependent on the success of viral cDNA synthesis in the reverse transcription step. A highly sensitive and reliable *in vitro* reverse transcription method has been developed in this work and has been used to detect cell-free HIV-RNA and HCV-RNA directly from the plasma or serum of seropositive patients and factor VIII concentrates (Zhang *et al.*, 1991). This method uses a virus-specific primer for initiation of cDNA synthesis from the viral RNA template. Others investigators have used random priming with hexameric oligonucleotides (Garson *et al.*, 1990). There has been no data so far that formally compared effectiveness of these two methods. In this work, viral RNA was firstly incubated with RNase-free DNase (BCL) at 37^{0}C for 20 minutes in a 10µl volume of DNase reaction buffer (50mM Tris-Cl pH7.5; 10mM $MgCl_2$; 4mM DTT; 10 units RNAsin and 15 units of RNAse-free DNase) to remove any possible HIV DNA contamination. The sample was then incubated at 80^{0}C for 10 minutes to terminate the reaction. cDNA synthesis was carried out by adding an equal volume of reverse transcriptase reaction buffer (50mM Tris-Cl pH8.0; 5mM $MgCl_2$; 5mM DTT; 50mM

KCl; 0.05ug/µl BSA; 600µM of each dGTP; dATP; dTTP; dCTP; 20% DMSO; 1.5µM outer anti-sense primer; 10 units RNAsin [Promega] and 10 units AMV reverse-transcriptase [Promega]) to the DNase-digested HIV RNA sample and incubating at 42^{0}C for 30 minutes.

2.2.6 Measurement of Reverse Transcriptase (RT) Reaction Efficiency by Plasmid RNA Transcription Construct pSV2gpt, containing a mouse promoter region and coding region of the herpes simplex virus type 1 thymidine kinase gene (HSV tk) was obtained from R. Al-Shawi (Division of Biological Sciences, University of Edinburgh). RNA was transcribed *in vitro* from 100ng of the construct plasmid DNA at 37^{0}C for 1 hour (20µl volumes; of 4mM Tris-HCl pH8.0; 8mM $MgCl_2$; 2mM spermidine; 50mM NaCl; 0.01M DTT; 0.4mM of each rATP, rTTP, rCTP and rGTP; 30 units of RNAsin; 100ng/µl BSA; 10 units of T7 RNA polymerase). The concentration of pSV2gpt RNA transcripts used for subsequent cDNA synthesis and of HSV tk plasmid DNA used for quantitative comparison was estimated by spectrophotometry at 260nm. The ration of optical density at 260nm and 280nm was also estimated to check the purity of synthesised RNA. Two fold serial titrations of tk cDNA after reverse transcription with anti-sense primer and of HSV tk plasmid DNA were made prior to PCR amplification with HSV tk specific primers. 25 cycles were employed and the products of PCR were analyzed by agarose gel electrophoresis and ethidium bromide staining. The amount of cDNA was estimated by reference to a dilution series of HSV tk DNA after amplification with the same primers. This was then compared with the number of RNA sequences

from which the cDNA was made.

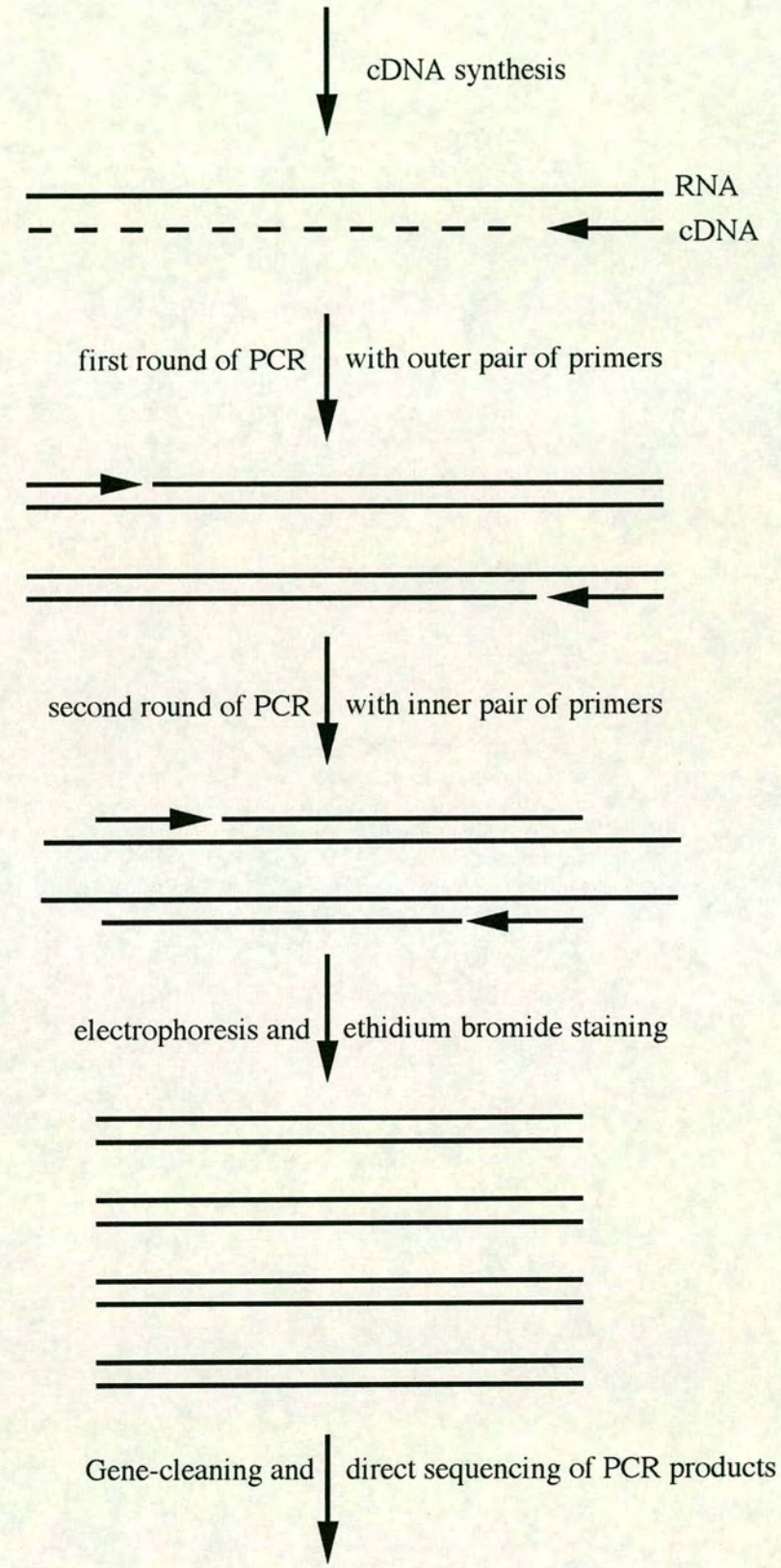
2.2.7 Detection and Quantification of HIV DNA and cDNA by Double Polymerase Chain Reaction Using Nested Primers Polymerase chain reaction is powerful *in vitro* method for the enzymatic synthesis of specific DNA sequences, using two oligonucleotide primers that hybridize to opposite strands and flank the region of interest in the target DNA. A repetitive series of cycles involving template denaturation, primer and template annealing, and the extension of the annealed primers by DNA polymerase results in the exponential accumulation of a specific fragment whose termini are defined by the 5' ends of the primers (see Figure 2.2). The double PCR method, which has been developed in our lab and has been used to detect and quantify HIV-1, was carried out in two consecutive steps. After the first PCR reaction finished, a small amount (usually 1μl) of the PCR product was transferred to a new tube which contained exactly the same solution as before except the primers which lie in the inner part of first pair of primers. This net effect is an overall increase in sensitivity and specificity of the PCR reaction, producing a 10,000-fold increase in amplification over that achieved by the only one pair of primers (Simmonds *et al.,* 1990*a*). Furthermore, the double PCR method is the only approach so far which is sensitive enough to amplify a single copy of the target sequences, and therefore can be used as a means of quantification of HIV DNA and RNA directly from patient materials (Simmonds *et al.,* 1990*a*). In experimental

details, the template DNA or cDNA was firstly limit diluted[2] and then presented in 20 to 50µl volume of 67mM Tris hydrochloride pH8.8; 16.7mM ammonium sulphate; 6.7mM $MgCl_2$, 10mM 2-mercaptoethanol; 6.7µM EDTA, 3.3µM each of dGTP, dATP, dTTP and dCTP, 170ug of bovine serum albumin per ml, 10% dimethyl sulphoxide, 0.5µM of each of the outer nested primers, and 20 units/ml of Taq polymerase [Cetus]. Twenty-five heat cycles were used, each consisting of 25 seconds at 94°C, 35 seconds at 50°C, and 2.5 minutes at 68°C. One microlitre of the product was then transferred to a second tube containing the same buffer as before but with the inner pair of nested primers. A further 25 cycles were carried out under the same conditions. The products of the second reaction (20µl) were analyzed by agarose gel electrophoresis and ethidium bromide staining. The outline of the whole procedure start from cDNA synthesis to agarose gel electrophoresis of PCR products is illustrated in Figure 2.2.

2.2.8 Agarose Gel Electrophoresis of PCR Products Six grams of low melting agarose (IBI) was dissolved in 300ml (2%) of 1 x TBE (0.089M Tris-borate and 0.089M boric acid, 0.01M EDTA; pH8.2 - 8.9) buffer at 90°C. After the gel solution had cooled to around 45°C, ethidium bromide was added (final concentration 0.5ug/ml), mixed and the gel was poured onto a pre-levelled 20cm x 20cm electrophoresis gel plate (Pharmacia) and left to polymerize at 4°C. The electrophoresis was carried out at 150 volts for about 10-30 minutes depending on

[2] Dilution at which less than 25% of the subsequent PCR amplification products are positive.

Figure 2.2 Outline of cDNA synthesis, double PCR amplification and direct sequencing of PCR products. Both first and second rounds of PCR amplification consist of 25 repeated cycles of template denaturing, template and primer annealing and new strand synthesizing (for more details, see section 2.2.7).


cDNA synthesis
RNA
cDNA
first round of PCR
with outer pair of primers
second round of PCR
with inner pair of primers
electrophoresis and
ethidium bromide staining
Gene-cleaning and
direct sequencing of PCR products
sequence reading, alignment and phylogenetic analysis

the length of the PCR products. The gel was then observed under the UV light and, if necessary, a photograph was taken using a Polaroid camera.

2.2.9 Analysis of Length Variation of Double PCR Products by Polyacrylamide Gel Electrophoresis The length variation of the double PCR product (either amplified from a single molecules or from undiluted DNA and cDNA) could be clearly resolved by polyacrylamide gel electrophoresis due to its high resolution. In this case, the second PCR reaction was performed exactly the same as before (in 20µl reaction solution) but with the half concentration of dNTPs (1.5µM) and 0.25µl of additional α-^{35}S-dATP (1000ci/mM, Amersham). One microliter of the PCR product was mixed together with 4µl of loading buffer (95% Formamide, 20mM EDTA, 0.1% Bromophenol Blue and 0.1% Xylene Cyanol FF); 2µl of distilled water, and then heated at 95^{0}C for 2 minutes before loaded on a 8% denaturing polyacrylamide gel (for 150ml sequencing gel mix: 75g urea, 30ml 40% acrylamide/bisacarylamide, 15ml 10 x TBE pH8.2 - 8.9, 0.15g ammonium persulfate, H_2O to 150ml) . The electrophoresis was proceeded at 75 watts for around 3 hours and the gel was fixed, dried and exposed to X-ray film (Kodak XAR-2 or AGFA CURIX) as described below (section 2.2.11).

2.2.10 Direct Sequencing of Double PCR Products The PCR fragment for subsequent sequencing reaction was amplified in a larger volume (50µl). The PCR product was then purified by treating the reaction mixture with Gene-Clean Kit (Bio 101, Inc.) and eluting with 1 x TE buffer (10mM Tris.Cl pH7.4, 1mM EDTA, pH8.0)

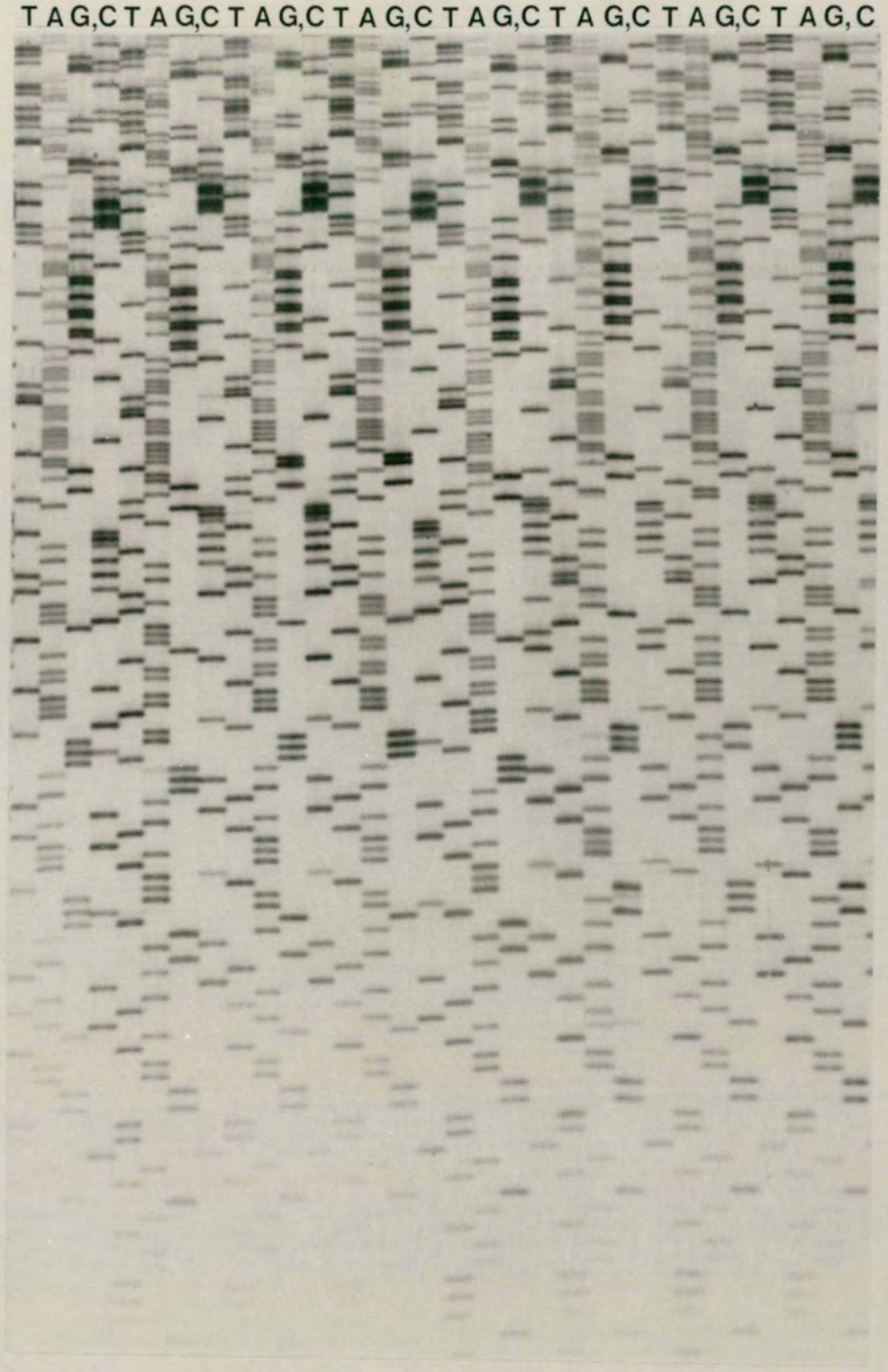
in order to remove the non-incorporated nucleotide triphosphates and primers. The sequencing was performed using the Sequenase Kit from United States Biochemical Corporation (USB), following a modification of the protocol of Winship (Winship, 1989). In this protocol, the use of culture grade DMSO at 10% in the sequencing reactions is recommended to enhance the intensity of the signal and to reduce background. Briefly, the purified PCR product (100-200ng) was mix with annealing mix (10% DMSO; 200mM Tris.HCl pH7.5; 100mM $MgCl_2$; 250mM NaCl; 10ng primer) and was boiled for 3 minutes to denature the template. This mixture was then immediately put on ice for 10 minutes to minimise template renaturation. The cooled annealed template/primer was mixed together with extension mix (0.025M DTT, 1 in 20 diluted labelling mix (7.5µl dGTP, 7.5µl dCTP, 7.5µl dTTP), α-^{35}S-dATP and 2 units Sequenase) on ice and aliquoted into appropriate pre-warmed (37^0C) termination mix (80µM dNTP [dGTP, dATP, dCTP, dTTP]; 8µM ddNTP [ddGTP, ddATP, ddCTP, ddTTP, respectively]; 50mM NaCl and 10% DMSO). The termination was stopped after 5 minutes incubation at 37^0C by adding 4µl of Stop solution (95% Formamide, 20mM EDTA, 0.1% Bromophenol Blue and 0.1% Xylene Cyanol FF). The resulting sample was heated up to 95^0C for 2 minutes before loaded on a 8% denaturing polyacrylamide gel.

2.2.11 Analysis of Nucleotide Sequence of Double PCR Product by Polyacrylamide Gel Electrophoresis Eight percent wedge sequencing gels were used (0.4mm at the top to 1.2mm at the base) as they can give better resolution and allow more nucleotides to be read. For 150ml of 8% gel, 75g urea (IBI), 12g

acrylamide (BDH), 0.6g bis-acrylamide (BDH), 0.15g ammonium persulphate (Sigma) and 15ml 10 x TBE (pH 8.8 - 8.9) were dissolved in 50ml of distilled water. Twenty microlitres of TEMED (N,N,N'N'-Tetramethylethylenediamine, Sigma) was added before pouring the gel between thoroughly ethanol cleaned wrapped glass plates. Just before loading, heat denaturing the sequencing sample at 95^{0}C for 2 minutes, and then load onto a pre-run (10 minutes) polyacrylamide gel. Electrophoresis was applied at 75 watts until the bromophenol blue reached the bottom of the gel (equivalent to 45 base pairs away from the 5' of primer). Following electrophoresis, the gel was fixed by soaking in 1 litre of 5% acetic acid and 5% methanol for 15 minutes, followed by a second wash in another litre of fixative for another 10 minutes. The gel was then dried on a gel dryer (Model 583 Gel Dryer, *BIO-RAD*) for 2-3 hours at 80^{0}C, and exposed to Xay film (Kodak XAR-2 or AGFA CURIX) in a cassette for about 15-24 hours. Films were developed in a automatic X-ray film processor (X-ograph X1). As one example, Figure 2.3 shows a developed film which had been exposed to a sequencing gel for 24 hours. The sequences on the film are part of the V4 region of HIV-1 in patient 82 (for more sequence detail, see Chapter 4).

2.2.12 Sequence Alignment, Determination of Nucleotide Distance and Phylogenetic Tree Construction The nucleotide sequences obtained from the developed films were collected and aligned using PILEUP and LINEUP programs on the University of Wisconsin GCG package (Devereux *et al.*, 1984). The PHYLIP package (version 3.4) of programs provided by Dr. J. Felsenstein (Department of

Figure 2.3 Direct nucleotide sequencing of double-stranded PCR products. The PCR products were derived from single molecules of HIV cDNA, reverse transcribed from viral RNA and limit diluted prior to PCR amplification. Sequencing reaction was carried out using HIV primer (x) (see section 2.1.4). Sequences shown are part of the V4 region of HIV-1 in patient 82 (for more sequence details, see Chapter 4).


T A G, C T A G, C T A G, C T A G, C T A G, C T A G, C T A G, C T A G, C

Genetics, University of Washington, Seattle) was used to perform phylogenetic analysis (Felsenstein, 1988). First, a matrix of nucleotide sequence distances were estimated using the program DNADIST. Average sequence diversities both within and between samples could be calculated on the basis of this matrix. Phylogenetic trees were constructed in two ways; first, by clustering the matrix of nucleotide sequence distances under the neighbor-joining algorithm of Saita and Nei (PHYLIP program NEIGHBOR) and second, by the more complex (and probably more reliable) maximum likelihood method (PHYLIP program DNAML). The evolutionary models underlying these different methods of phylogenetic reconstruction are outlined in the documentation to PHYLIP (Felsenstein, 1988) (see section 1.2).

2.2.13 Slot blot of Viral RNA A piece of Hybond^{TM}-N membrane (11cm long and 3.5cm wide, Amersham) was soaked briefly in distilled water and then in 20 x SSC (3M NaCl [175g/l], 0.3M Na_3citrate.$2H_2O$ [88g/l], adjust Ph to 7.0 with 1M HCl) for 30 minutes. Meanwhile, clean the manifold carefully with 0.1M NaOH and then rinse it well with distilled water. The wet Hybond^{TM}-N membrane was put onto the manifold avoiding any air bubbles between the manifold and the membrane. Clamp the parts of the manifold together, and connect the vacuum unit to a vacuum pump. The slots were washed twice with 10 x SSC by applying gentle suction before loading the samples. The pre-treated (in 50% formamide; 7% formaldehyde and 1 x SSC solution at 68^{0}C for 10 minutes and cooled on ice) HIV viral RNA (HIV_{HXB2}, HIV_{RF} and $HIV_{HTLV\text{-}IIIB}$) for cDNA synthesis was serially diluted (in DEPC-treated sterilized water) in two-fold steps and were deposited onto a Hybond^{TM}-N

hybridization transfer membrane by vacuum suction on a 24 slots HYBRI-SLOT™ MANIFOLD (142.2cm long and 6.3cm wide, BRL). A dilution series of known amounts of the pre-treated (95°C for 10 minutes and chilled on ice) HIVBH01R.3 plasmid DNA were also slot blotted onto the same membrane as a quantification standard. After all the samples have passed through the filter, the slots were rinsed again twice with 10 x SSC. The Hybond™-N membrane was removed from the manifold, and allow it to dry completely at room temperature followed by UV fixation for 3 minutes. The membrane, at this stage, is ready for hybridization.

2.2.14 Probe Synthesis The selection of an appropriate combination of label and labelling method for a particular experiment depends mainly on the level of sensitivity and resolution required. For most filter hybridization applications, sensitivity is considered to be of greater importance. Accordingly, for maximum sensitivity in filter hybridization, phosphorus-32 is the most widely used radiolabel as it is available at high specific activity and can be detected with a high degree of efficiency. Several methods are available for making probes. End-labelling with polynucleotide kinase has been employed with short oligonucleotides, whereas longer DNA fragment have been labelled by nick translation and random priming methods.

2.2.14.1 Random Priming Method In this work, random primer (Multiprimer) labelling method was employed which was firstly suggested by Feinberg and Vogelstein (Feinberg *et al.*, 1983). This approach utilizes the ability of DNA polymerase I to synthesize a new DNA strand complementary to a template strand, starting from a free 3'-hydroxyl. In this case the latter is provided by random

hexanucleotides derived either from DNase I digestion of calf thymus DNA or by oligonucleotide synthesis. The 'Klenow' fragment of DNA polymerase I is used because it lacks the 5'-3' exonuclease activity which would otherwise degrade the primers. The absence of the 5'-3' exonuclease also ensures that incorporated nucleotides are not subsequently removed as monophosphates. Random primer labelling reactions can be carried out at room temperature or at 37^{0}C for about 30 minutes. Linear single-stranded or denatured double-stranded DNA molecules are usually used as substrates. Covalently-closed circular DNA can also be used, but slightly lower incorporations are obtained. In this work, the HIVBH10.R3 plasmid DNA (50ng in DEPC-treated sterilized water) was denatured by boiling for 3 minutes and chilled on ice. The mixture of 6ul of OLB[3], 1.2µl of 10mg/ml of bovine serum albumin (BSA), 5µl of [α-^{32}P] dATP and 1 unit of Klenow fragment (BCL) were then added to denatured DNA template and incubated at 37^{0}C for 30-60 minutes. This procedure routinely allows to obtain specific activities of 1 x 10^9 dpm/ug.

2.2.14.2 Removal of Unincorporated Nucleotides and Proteins In general, the denatured labelled DNA can be used directly as a hybridisation probe without stopping the reaction or removing unincorporated label. Sometimes, when only very small amount of target RNA or DNA molecules are present on the hybridization membrane, further purification of labelled probe is necessary in order to get rid of any possible interference either from unincorporated nucleotides and from protein

[3] OLB is made up by 50ul solution A (1.25M Tris-HCl pH8.0; 0.125M $MgCl_2$; 25mM-mercaptoethanol; 0.5mM each of dGTP, dTTP, dCTP), 125ul solution B (2M HEPES buffer titrated to pH6.6 with NaOH) and 75ul solution C (random hexanucleotides OD_{260}= 90 units/ml in 3mM Tris-HCl and 0.2 mM EDTA, pH7.0) (Feinberg and Vogelstein *et al.*, 1984).

debris. In the latter case, the labelling reaction is stopped by adding 1µl of 0.5M EDTA solution followed by phenol and chloroform extraction. The aqueous phase, which contains DNA probe, is precipitated with 40µl of 5M ammonium acetate $(NH_4)_2Ac$ and 200µl of 100% ethanol at -70°C for about 30 minutes to an hour. The precipitated DNA was collected by centrifugation at 4°C for 10 minutes and finally, the pellet was washed with 70% ethanol and resuspended in 100µl of distilled water. This relatively clean probe was boiled for 5 minutes before it was put into use in hybridization process.

2.2.15 Hybridization with Homologous DNA Probe The hybridization reaction is influenced by a number of factors, some related to the nature of the probe used, others to the general conditions under which the reaction is carried out. In general, to maximise the rate of annealing of the probe with its target, hybridizations are usually carried out in the presence of 10% dextran sulphate or 10% polyethylene glycol, in a solution of high ionic strength at a temperature that is 20-25°C below the melting temperature (T_m[4]). This is especially true when oligonucleotide probes are used. To minimize background problems, pre-hybridization with a blocking agent (5 x Denhardt's reagent[5], 0.5% SDS, and 100ug/ml denatured, fragmented salmon sperm DNA) should be employed and the washing conditions should be as stringent

---

[4] Tm (melting temperature): the temperature at which hybrids of a particular probe and its complementary sequence are 50% dissociated or denatured. Tm is mainly dependent on ionic strength, base composition and denaturing agents.

[5] (0.1% BAS-Pentax Fraction V; 0.1% Ficoll and 0.1% polyvinylpyrollidone)

as possible. Pre-hybridization and hybridization with homologous DNA probe was performed in rather stringent conditions (0.5M Na_2HPO_4/NaH_2PO_4, pH 7.2;, 1mM EDTA and 7% SDS at 65^{0}C overnight) to reduce non-specific annealing. Non-specifically bound nucleotides were removed by washing the membrane twice with pre-warmed (57^{0}C) washing solution (40mM sodium phosphate Na_2HPO_4/NaH_2PO_4, pH 7.2;, 1mM EDTA and 1% SDS) for 20 minutes each. Filters were then ready for autoradiography.

2.2.16 Autoradiography Damp filter was placed on a sheet of Saran Wrap and exposed to X-ray film (Kodak XAR-2 or AGFA CURIX) in a cassette. The exposure time is normally 16-24 hours at -70^{0}C with an intensifying screen.

CHAPTER 3

DETECTION AND QUANTITATION OF HIV AND HCV BY A NEWLY DEVELOPED RNA PCR METHOD

3.1 SUMMARY

3.2 INTRODUCTION

3.3 RESULTS

3.3.1 Optimizing the Reaction Conditions of Reverse Transcription

3.3.2 Reverse Transcriptase Reaction Efficiency from Plasmid RNA Transcription

3.3.3 Reverse Transcriptase Reaction Efficiency Using HIV RNA

3.3.4 Quantitation of HIV Particles in the Plasma or Serum of Seropositive Individuals

3.3.5 Detection and Sequencing of HIV-1 in Factor VIII Concentrates

3.3.6 Detection and Quantitation of HCV RNA in Factor VIII Concentrates

3.4 DISCUSSION

3.4.1 Detection and Quantitation of Cell-free HIV RNA in Plasma of HIV-1 Seropositive Individuals

3.4.2 Detection of HIV RNA from Factor VIII Concentrates

3.4.3 Detection and Quantitation of HCV RNA from Factor VIII Concentrates

3.4.4 Proposed Substitution for AZT Resistance Was a Pre-existing Polymorphism

3.1 Summary

A highly sensitive and reliable RNA polymerase chain reaction method has been developed which can be used to detect, quantify and sequence cell-free HIV RNA directly from the plasma or serum of seropositive individuals. Plasma from 10 out of 12 haemophiliacs tested was found to contain detectable levels of HIV-1 RNA [geometric mean value: 1.2×10^3 copies for CDC (Centres for Disease Control) group II patients, 5.5×10^3 copies for CDC group IV patients]. The presence of cell-free circulating virus in both symptomatic and asymptomatic individuals suggests that viral replication continues throughout the course of infection. The same procedure has also been applied to detect, quantify and sequence HIV-1 RNA in two batches of unheated commercial factor VIII concentrates distributed in 1981 and 1983. The sequences obtained revealed a closer relationship to North American than to African variants of HIV-1. The amounts of HCV RNA present in factor VIII concentrates are substantially higher than those of HIV RNA.

3.2 Introduction

By means of the polymerase chain reaction (PCR), HIV provirus (DNA) can not only be detected (Ou *et al.*, 1988; Saiki *et al.*, 1988), but also accurately quantified directly in peripheral blood mononuclear cells (PBMCs) of HIV-1 positive individuals (Simmonds *et al.*, 1990*a*; Oka *et al.*, 1990). However, the detection of HIV DNA in PBMCs does not indicate whether such cells are expressing viral RNA sequences or whether free virus is present in plasma or other body fluids. Recently, several investigators have coupled a reverse transcriptase (RT) reaction step to the

polymerase chain reaction (RNA PCR) and have successfully detected HIV RNA both in cultured HIV-1 infected cell lines and in PBMCs from HIV-1 seropositive subjects (Hart *et al.*, 1988; Byrne *et al.*, 1988). Unfortunately these reports have not included an assessment of the sensitivity of the methods used, either in terms of the efficiency of the reverse transcriptase reaction or of minimum number of HIV RNA molecules required to produce a positive PCR signal. Therefore, these studies, though capable of measuring relative quantities, have not determined the absolute amounts of RNA present in the study subjects. In this work, a highly sensitive and quantitative RNA PCR assay has been developed. After reverse transcription, cDNA was amplified in two sequential PCRs. As the nested PCR can detect single molecules of target DNA sequence (Simmonds *et al.*, 1990*a*), quantitation of HIV-specific cDNA, and by implication of HIV RNA sequences present in the original samples, can be achieved as described previously for provirus quantification in PBMCs (Simmonds *et al.*, 1990*a*). The absolute quantification of RNA requires knowledge of the efficiency of the reverse transcriptase reaction. This was obtained by estimating the yield of cDNA from known amounts of specific RNA sequences after reverse transcription. These measurements were made both with a transcribed HSV tk (herpes simplex virus type 1 thymidine kinase gene) RNA template amplified with HSV tk specific primers, and HIV RNA template amplified with *env*-gene specific nested primers. The efficiency of the RNA PCR was studied under different reaction conditions and the optimum has been established. The method was then used to detect and quantify the amount of HIV RNA present in plasma of HIV-infected individuals. The amount of circulating virus in infected patients was

compared with clinical status, CD4+ lymphocyte counts and the amount of virus in PBMCs. The presence of HIV-1 RNA and HCV (hepatitis C virus) RNA sequences in 8 batches of unheated factor VIII concentrate distributed between 1981 and 1984 has also been investigated. The levels of HIV and HCV RNA in the factor VIII was then compared. HIV RNA was detected in two and HCV in six out of eight batches of commercial factor VIII concentrates distributed in 1981 and 1983 (including the two positives for HIV RNA). The identity of the HIV RNA was confirmed by nucleotide sequencing the PCR product. Sequences obtained in the *pol* and *env* regions from these concentrates provide some information on the geographical origins of the infected blood donors.

3.3 Results

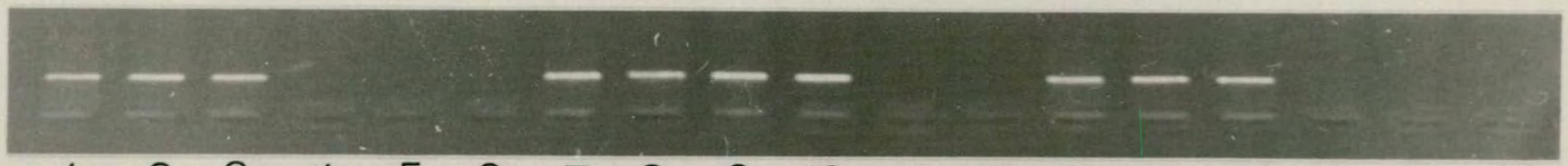
3.3.1 Optimizing the Reaction Conditions of Reverse Transcription The known amount of HIV RNA, measured by its optical density at 260nm, was reverse transcribed under the different reaction condition. The cDNA product was then serially diluted (10-fold) before the subsequent double PCR reaction. The last dilution at which amplified DNA was detectable by agarose gel electrophoresis and ethidium bromide staining was recorded and used subsequently to find out the better reaction condition. Firstly, the reverse transcription reaction was carried out at different temperature, 37°C or 42°C and for 3 different periods (10 minutes., 20 min. and 30 min.). The dilution limit of the reaction which was carried out at 42°C for 30 minutes was at least 10 fold lower than for reactions carried out under the other

conditions (Fig. 3.1, lanes 19 to 24). The longer period of time (one or two hours) were applied to the same reaction condition (42^{0}C) and that seems to not increase the yield of cDNA products (data not shown). Interestingly , the presence of carrier RNA (sheep fibroblast total RNA) in 0.5 ug to 1.5 ug quantities in the reaction buffer can greatly enhanced the reverse transcription efficiency with both *gag* and *env* anti-sense primers (Fig. 3.2). However, an inhibitory effect of carrier RNA was also observed if a large amount was present in the reaction (Fig. 3.2). Therefore, the 1ug quantity of carrier RNA is included in all of the subsequent reverse transcription reactions carried out in this work. The presence of DMSO and BSA in the buffer seems also to increase the yield of the RT reaction (data not shown), and therefore were included in the reaction mix (Methods section 2.2.5).

3.3.2 Reverse Transcriptase Reaction Efficiency from Plasmid RNA Transcription RNA was transcribed by T7 RNA polymerase from the construct pSV2gpt, containing a mouse promoter region and coding region of the HSV type 1 thymidine kinase gene (HSV tk) (see section 2.2.6), and quantified by spectrophotometry at 260nm. cDNA synthesized from 8ng (nanogram, 10^{-9} gram) of HSV tk transcript was serially diluted prior to single PCR amplification with tk-specific primers. A dilution containing cDNA synthesized from 16 fg RNA gave a positive result and the next dilution containing cDNA from 8 fg was negative (Fig. 3.3, lanes 1 to 11). Using the estimated molecular mass of the RNA transcript (2400 base x 330g/mol per base), the minimum detectable amount of cDNA corresponded to an input of 1.2 x 10^4 copies of RNA. This result was compared with the results

Figure 3.1 Difference in Efficiency of the Reverse Transcription Reaction under Various Reaction Conditions. cDNA were ten-fold serially diluted prior to PCR amplification. Lanes 1 to 6, and 7 to 12: cDNAs were synthesized at 37^{0}C for 20 minutes and 30 minutes, respectively. Lanes 13 to 18, and 19 to 24: cDNA were synthesized at 42^{0}C for 20 minutes and 30 minutes, respectively. PCR amplification of cDNA were carried out in two consecutive steps using nested primers (outer primer pair are [w], [z] and inner ones are [x] and [y]). The products of the second reaction were analyzed by agarose gel electrophoresis and ethidium bromide staining (see section 2.2.7).

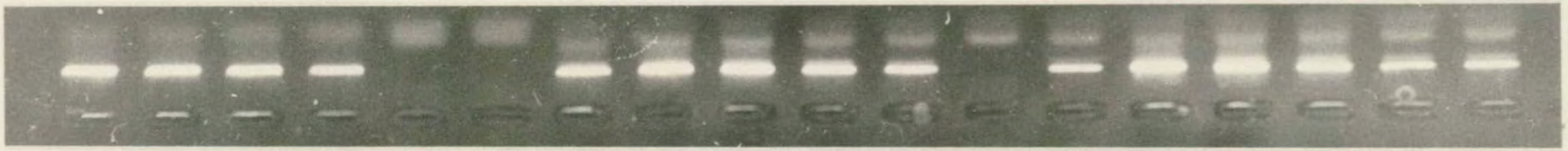
19 20 21 22 23 24.



1 2 3 4 5 6. 7 8 9 10 11 12. 13 14 15 16 17 18.

Figure 3.2 Difference in HIV RNA Detection Efficiency with and without Carrier RNA. cDNA were ten-fold serially diluted before double PCR amplification. Lanes 1 to 6: no carrier RNA presented in the reverse transcription reaction mix. Lanes 7 to 12: 0.5μg carrier RNA was added to the reaction mix. Lanes 13 to 18, 19 to 24, 25 to 30, 1μg, 1.5μg and 3μg of carrier RNA were present during reverse transcription. PCR amplification and detection of PCR products were carried out as described in the Legend to Figure 3.1.

19 20 21 22 23 24, 25 26 27 28 29 30.



1 2 3 4 5 6, 7 8 9 10 11 12, 13 14 15 16 17 18,

of single PCR amplifying a dilution series of cloned tk DNA. The cut-off point was 5.12 fg (femtogram, 10^{-15} gram) for the tk plasmid (Fig. 3.3, lanes 12 to 22). Using the known molecular mass of the tk plasmid DNA (7270 bp x 660g/mol per bp), this figure corresponded to 630 molecules of tk plasmid, or 1260 copies of target sequences (630 x 2 for double-stranded DNA). The efficiency of the RT reaction, in terms of the number of molecules of amplifiable cDNA synthesized from the RNA template, is the ratio of the two figures obtained above (1260/12,000), or approximately 10%.

3.3.3 Reverse Transcriptase Reaction Efficiency Using HIV RNA Six HIV-1 RNA samples were extracted from culture supernatant of C8166 cells infected with HIV_{RF}, $HIV_{HTLV\text{-}IIIB}$ and HIV_{HXB2} (see section 2.2.4). The viral RNA was slot blotted and quantified by hybridization with HIVBH10.R3 plasmid probe, in comparison with a dilution series of known amount of HIVBH10.R3 DNA (Figure 3.4). Reverse transcription reactions containing 19.8, 9.9, 7.0 and 0.6 pg (picogram, 10^{-12} gram) HIV_{RF} viral RNA, 3.0 pg $HIV_{HTLV\text{-}IIIB}$ viral RNA and 1.5 pg HIV_{HXB2} viral RNA were then carried out with an HIV V4-V5 outer anti-sense primer (primer z; see section 2.1.4 and Figure 2.1). The number of copies of cDNA in each sample after reverse transcription was estimated by limiting dilution prior to PCR amplification with nested primers (outer pair [w] and [z] and inner pair [x] and [y], see section 2.1.4 and Figure 2.1).

The frequency of positive reactions at limiting dilution was used to calculate the molecular concentration of cDNA using a Poisson correction for positive reactions

Figure 3.3 Measurement of Reverse Transcriptase Reaction Efficiency from Plasmid RNA Transcription. Lanes 1 to 11: two-fold serial dilution of cDNA synthesized from 8ng of HSV tk transcript before single step PCR amplification with a pair of HSV tk-specific primers. Lanes 12 to 22: two-fold serial dilution of known amount of plasmid tk DNA, prior to single PCR amplification. PCR amplification and detection of PCR products were carried out as described in the Legend to Figure 3.1.

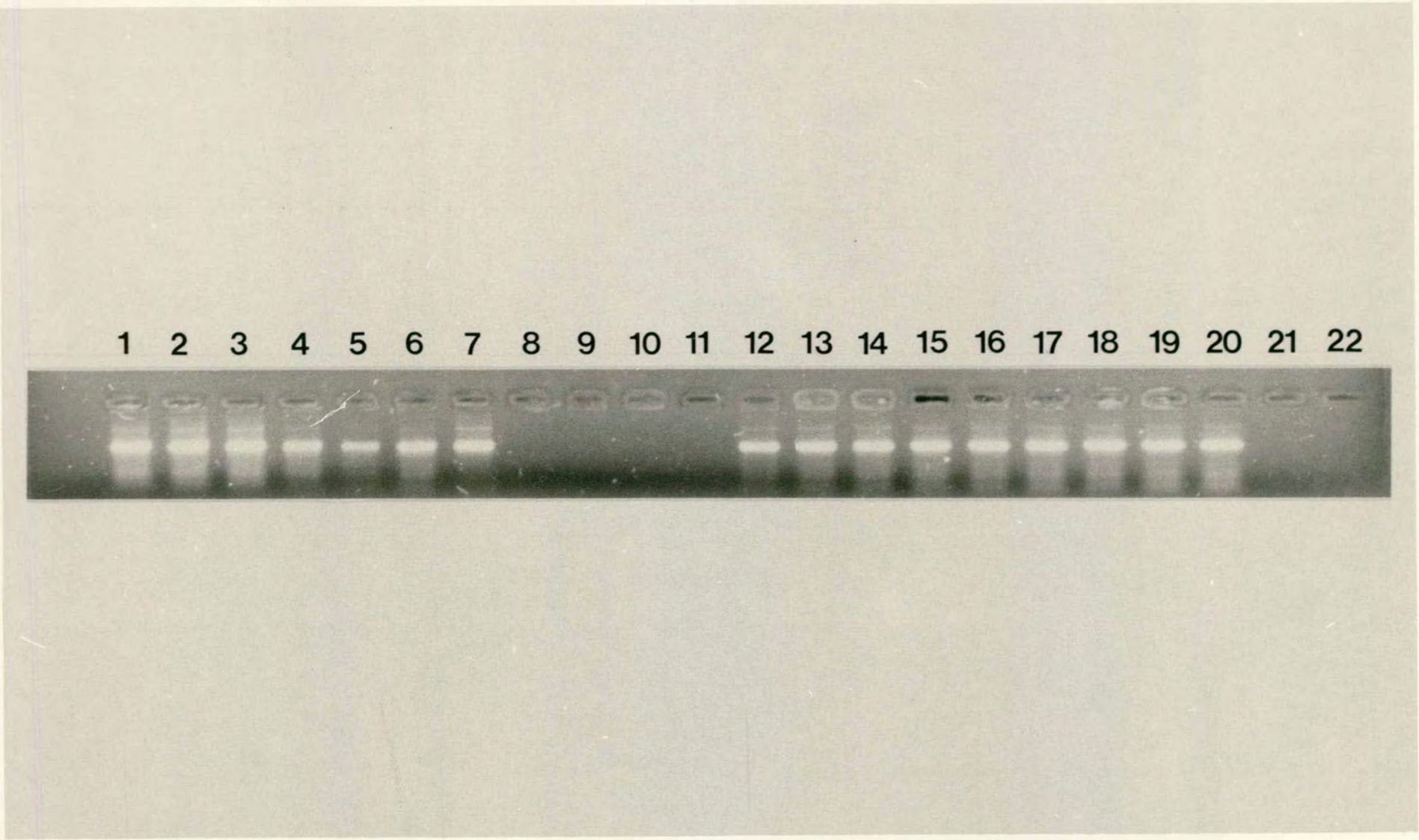
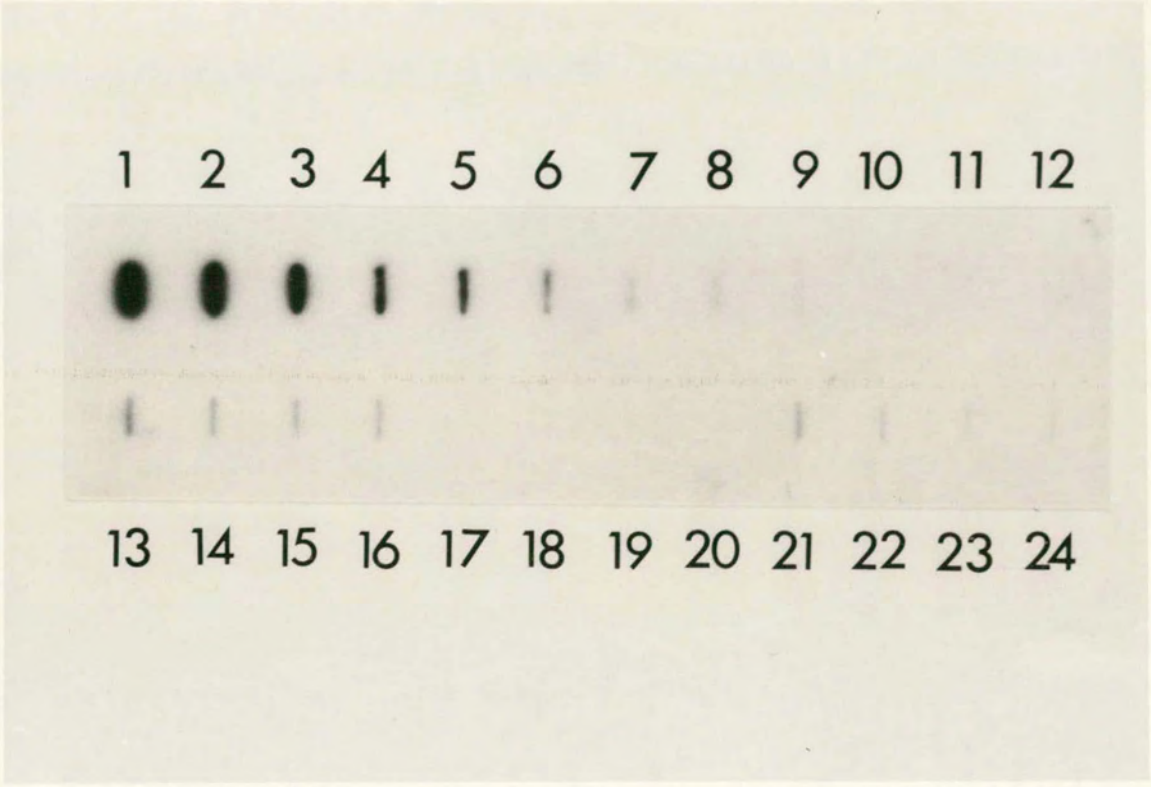

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22

Figure 3.4 Quantitation of viral RNA by slot blot and hybridization method (for details see sections 2.2.13 and 2.2.15). Lanes 1-12, two-fold serial dilutions of known amount of HIVBH10.R3 DNA (3ng, 1.5ng ... 1pg in slot 1, 2...12, respectively). Lanes 13-16, two-fold serial dilution of viral RNA samples from $HIV_{HTLV\text{-}IIIB}$; lanes 17-20, from HIV_{HXB2}; lanes 21-24, from HIV_{RF}.


1 2 3 4 5 6 7 8 9 10 11 12
13 14 15 16 17 18 19 20 21 22 23 24

which contain more than one template molecule. [The mean number of template molecules per reaction (m) is equal to $-\ln(f_0)$, where f_0 is the fraction of negative reactions.] The molecular mass of one copy of HIV-1 RNA was 6.5 ag (attogram, 10^{-18} gram) (Simmonds *et al.*, 1990*a*) from which the RT reaction efficiency was calculated. In each case, 24 (or 20) replicates were tested at limiting dilution, and the estimates were reasonably reproducible (Table 3.1). For the first sample for instance, at a 1 in 78,125 dilution, there were two PCR positives out of 24 replicates. Using the Poisson formula to correct for multiple positives, the total number of molecules of HIV RNA detected at this dilution was estimated to be 2.1. As the molecular mass of single HIV RNA was 6.5 ag, the total amount of cDNA synthesized by the reverse transcription reaction was therefore equivalent to 1.06 pg ($2.1 \times 6.5 \times 10^{-6} \times 78{,}000$). The ratio of the amount of HIV RNA reverse transcribed and the initial amount of HIV RNA available for reverse transcription gave the RT efficiency (1.06/19.8 = 5.3%). The results of dilution and distribution of six independent cDNA reactions after double PCR amplification were shown in Table 3.1, along with the calculated RT efficiencies.

3.3.4 Quantitation of HIV Particles in the Plasma or Serum of Seropositive Individuals 200-500ul of plasma samples from 12 haemophiliacs was diluted in PBS (phosphate-buffered saline) and centrifuged at 20,000 x g for 2 hours at 4^0C to pellet virus (for details see section 2.2.4). The pellet was resuspended in guanidinium/phenol solution (Chomczynski *et al.*, 1987). RNA was purified as described in section 2.2.4. and finally dissolved in 20ul of DEPC-treated distilled

Table 3.1 Titration of HIV cDNA after RT reaction by dilution, distribution and double PCR amplification.

Sample (pg)	Amount of RNA in RT reaction (pg)	Dilution factor of cDNA (no. positives/no. tested)						Calculated amount of cDNA (pg)	% RT efficiency
		125	625	3125	15625	78125	390625		
RF1	19.8	ND	4/4	4/4	3/4	2/24	0/4	1.06	5.4
RF2	9.9	4/4	4/4	4/4	4/24	0/4	0/4	0.44	4.4
RF3	7.0	4/4	4/4	4/4	4/24	0/4	0/4	0.44	6.3
RF4	0.6	3/4	5/24	0/4	0/4	0/4	0/4	0.02	3.3
IIIB	3.0	4/4	3/4	1/4	2/20	0/4	ND	0.21	7.0
HXB2	1.5	2/4	2/4	4/20	0/4	0/4	ND	0.09	6.0

ND, not done

water. 7ul of this RNA solution was treated with RNAse-free DNAse first and then reverse transcribed in a 20ul volume of reverse transcription reaction mix (for details see section 2.2.5). 10ul of cDNA was then detected and quantified in a nested PCR reaction using primers z,w and x,y (see figure 2.1).

Plasma from 10 out of 12 haemophiliacs contained detectable levels of HIV-1 RNA. The threshold of detection can be calculated to be 228-571 (on average 400) copies of RNA per ml plasma, based on the volume of plasma from which RNA was extracted, the input volume of RNA in the reverse transcription reaction, the proportion of cDNA used in the first PCR reaction and the efficiency of reverse transcription (e.g. 1 x 10^3/{200-500ul[plasma used for RNA extraction] x 7/20[input volume of RNA] x 10/20[input of cDNA in subsequent PCR reaction] x 5% [reverse transcription efficiency} = 228-571 viral RNA per ml of plasma). The estimated concentration of virus particles in plasma ranged from 1 x 10^3 to 3 x 10^4 copies per ml in the positive samples (Table 3.2), with a geometric mean value of 1.2 x 10^3 copies for CDC group II patient, and 5.5 x 10^3 copies for CDC group IV patients. There was a significant positive association between the concentration of viral RNA in plasma and the proviral abundance in PBMC DNA. The correlation coefficient between the log-transformed RNA and DNA estimates is 0.74 ($0.01<P<0.02$), indicating that over 50% of the variance in DNA proviral abundance can be explained by the RNA concentrations (Dr. Andrew.J. Leigh Brown, personal communication). Thus the patients with the lowest proviral abundance (83 and 84 with an average of one provirus in 14,000 and 10,000 PBMC, respectively) had less than 200 copies of RNA per ml in their plasma. On the other hand in patient 82 and 87, relatively

Table 3.2 Comparison of the amount of circulating virus from patients' plasma with the amount of provirus in PBMCs, plasma p24 antigen concentration, CD4+ counts, and clinical status.

Patient	Virions per ml in plasma	Number of cells per single provirus	p24 antigen (pg/ml)	CD4+ lymphocytes (x 10^9/l)	Zidovudine treatment (months)	Disease stage (CDC)
p83	$<10^2$	14000	-	0.27	-	II
p84	$<10^2$	10000	-	0.05	-	II
p77	1.26×10^3	2500	-	0.07	14	IVA
p56	1.29×10^3	ND	15	0.51	-	II
p74	3.49×10^3	2000	-	0.38	10	IVC2
p95	3.50×10^3	455	-	0.06	-	IV
p28	3.97×10^3	2718	300	0.09	-	IV
p79	4.39×10^3	3300	63	0.21	15	IVC2
p70	6.12×10^3	ND	-	0.39	-	II
p82	8.53×10^3	700	53	0.65	-	II
p72	2.76×10^4	2720	-	0.33	-	IVC
p87	2.96×10^4	589	20	0.05	17	IVC/E

higher concentrations of virus (8.5×10^3 and 3.0×10^4 per ml plasma) were associated with high frequencies of provirus-bearing PBMC (one in 700 and one in 589 PBMC infected, respectively). However, no correlation was found, regardless of the stage of infection, between the amount of virus in plasma and the level of p24 antigen. Five patients who were negative for p24 antigen (70, 72, 74, 77 and 95) contained over 1.0×10^3 virus particles per ml plasma. Furthermore, no correlation between CD4+ lymphocyte depletion and amount of circulating virus was found in these individuals. For example, relatively normal CD4 counts were found in p82 despite containing 8.5×10^3 copies of viral RNA per ml. Conversely, p84 had low CD4 counts (0.05×10^9 per litre) yet no detectable circulating virus. Zidovudine treatment appeared to have had little long term effect on the level of circulating virus. Those on long-term treatment (p74, 77 and 79) contained comparable levels to the two untreated symptomatic individuals (p95 and 72).

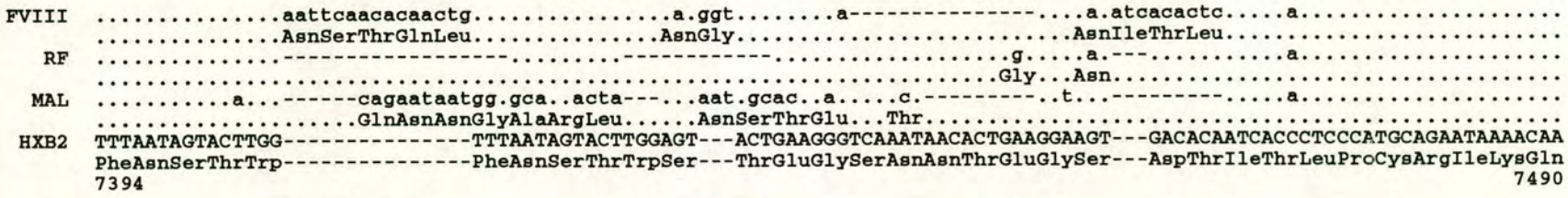
3.3.5 Detection and Sequencing of HIV-1 in Factor VIII Concentrates Eight batches of factor VIII concentrates, including both commercial and NHS (National Health Service)-produced material, were obtained from Dr. G. Kemball-Cook (National Institute for Biological Standards and Control [NIBSC]). All batches were unheated and prepared before the introduction of donor screening for anti-HIV antibodies. They were distributed in the United Kingdom between 1981 and 1984. All factor VIII concentrates were reconstituted according to the manufacture's instructions.

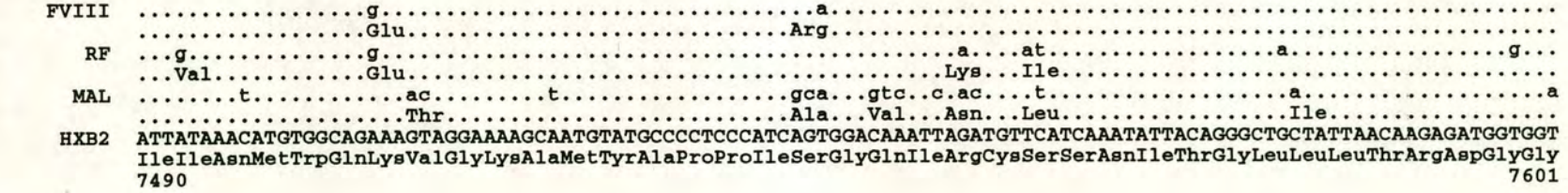
RNA was prepared from these batches of factor VIII by high-speed

centrifugation at 20,000 x g and solubilization of the virus pellets with guanidinium/phenol solution as described in section 2.2.4. One third of each RNA samples was reverse transcribed with HIV *pol* primer (l), or *env* primer (z) or HCV primer ED2, respectively. One-quarter of the cDNA was amplified by PCR with corresponding nested primers (see sections 2.2.5 and 2.2.7). Two factor VIII batches (both commercially derived) out of eight tested gave positive results for HIV-1 RNA; in one case with the *env* primers, the other with the *pol* primers. Single molecules of target cDNA were isolated by limiting dilution of the cDNA and were directly sequenced as described in section 2.2.10. In the *env* region, two HIV RNA sequences, obtained from batch no. 1, were identical in the V4 and C3 regions. The sequences were distinct from those of all published HIV isolates and from any HIV sequence obtained previously in our laboratory (Balfe *et al.*, 1990; Simmonds *et al.*, 1990*a*). This is particularly apparent in the V4 hypervariable region, which is clearly distinct from all published sequences (see Figure 3.5). The C3 region showed its 92% similarity with HIV_{HXB2}, 90% with HIV_{RF}, and 77% with HIV_{SF-2}. In the *pol* region, two HIV RNA sequences obtained from batch no. 8 were also identical and distinct from any published sequence (Figure 3.5). In this region, the sequence was 96% identical to HIV_{HXB2}, 94% with HIV_{RF} and HIV_{SF-2} and 92% with HIV_{Z6}. The amount of RNA present in this material was close to the threshold of sensitivity for the RNA PCR method used. Allowing 5% efficiency of reverse transcription with these primers, the calculated amount of HIV RNA in both batch of reconstituted factor VIII was only 2.5 copies per ml.

Figure 3.5 Comparison of sequences detected in factor VIII with those of known geographical variants of HIV-1. (a) Nucleotide and amino acid sequences of the V4 and C3 region (*env*) of RNA detected in factor VIII batch No. 1. (b) Sequences in two regions of *pol* of RNA from batch No. 8. The location of these sequences in the genome of the $HIV_{IIIB\text{-}HXB2}$ is indicated. Differences between the factor VIII sequences and those of the HIV_{RF} and HIV_{MAL} isolates from that of $HIV_{IIIB\text{-}HXB2}$ are shown in the body of the figure. Amino acid replacement from lysine to arginine at position 70, which conferring AZT resistance, is underlined in Pol 1 Region.

(a) V4/C3 Region





(b) Pol 1 Region

```
FVIII  ..........c.....................................a.g..........c.....................................g.....................
       ..........................................................Arg.........Ser.....................................Arg.....................
RF     ..........................................................................................................................g
       ..........................................................................................................................
MAL    ..........................................................c.....................ga..........g.........a...........t..t.............
       ..........................................................................Asn.....................................................
HXB2   GGGCCTGAAAATCCATACAATACTCCAGTATTTGCCATAAAGAAAAAAGACAGTACTAAATGGAGAAAATTAGTAGATTTCAGAGAACTTAATAAGAGAACTCAAGACTTCTGGGAAGTTCAA
       GlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGln
       2699                                                                                                                    2822
```

Pol 2 Region

```
FVIII  ..........................................................................................c.........a...
       ..........................................................................................Ser.............
RF     ..............a..............a.....................................................t...............
       ..............Lys............Glu.....................................................................
MAL    ........a..c.......c.a..........a.....c..a..c.........................g.....t...........a..a
       ..................ThrLys......Glu.....................................................................
HXB2   ATCTTAGAGCCTTTTAGAAAACAAAATCCAGACATAGTTATCTATCAATACATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAG
       IleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGln
       3047                                                                                    3140

FVIII  ....................a...................a.......t.......c.....................................
       ..........................................Lys......Phe.....................................
RF     .......t............a......g..........a.a......gt..........g.....g.........................t...
       ......Ile..................Glu.........Lys......Phe.....................................
MAL    ....................a..a...g.......a....aa......t...................g..............c.....t...
       ..............................Glu.........Lys......Phe.....................................
HXB2   CATAGAACAAAAATAGAGGAGCTGAGACAACATCTGTTGAGGTGGGGACTTACCACACCAGACAAAAAACATCAGAAAGAACCTCCATTCCTT
       HisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluProProPheLeu
       3140                                                                                    3233
```

3.3.6 Detection and Quantitation of HCV RNA in Factor VIII Concentrates Six out of eight batches of factor VIII concentrate tested contained detectable levels of HCV RNA. All the commercial-derived batches were RNA PCR positive for HCV, including the two positive for HIV RNA. However, two National Health Service-derived batches were negative for HCV RNA. The amounts of HCV RNA present in these factor VIII concentrates were substantially higher than that of HIV, in two batches that were quantified by limiting dilution method, between 30,000 and 100,000 copies of HCV RNA per ml were found (Simmonds *et al.*, 1990*b*).

3.4 Discussion

3.4.1 Detection and Quantitation of Cell-free HIV RNA in Plasma of HIV-1 Seropositive Individuals A highly sensitive and reliable RNA PCR method was developed which can be used to detect, quantify and subsequently sequencing directly from the patient plasma and serum without any *in vitro* cloning. By this means, the errors introduced by Taq polymerase during the PCR amplification process can be avoided. An efficiency of around 5% was obtained in the RT reaction of HIV-1 template and primers (spacing 480 bp). The overall efficiency of the procedure declined with wider primer spacing; amplification of cDNA using primer pairs separated by 858 bp gave an efficiency of 1.8% (data not shown).

All plasma samples used in these experiments were obtained during 1988 and 1989 from HIV-seropositive haemophiliacs who were infected in 1984 (Ludlam *et al.*,

1985). Five were asymptomatic and seven had AIDS or AIDS-related complex. Of the five asymptomatic (CDC group II) patients, three were RNA PCR-positive, while all seven CDC group IV patients were positive for RNA PCR (Table 3.2). There was no correlation between RNA titre and presence of circulating p24 antigen. All samples that were p24 antigen-positive were positive in the RNA PCR, however, high level of viral RNA sequences were also found in some p24 antigen-negative plasma samples (p70, 72, 74 and 77) while similar or lower amounts of circulating RNA have been found in other plasma samples that were antigen-positive (Table 3.2).

Previous studies have shown that p24 antigen present in patient plasma is readily detectable at certain stage of HIV infection. Antigenaemia is detectable for several weeks on primary infection with HIV (Goudsmit *et al.*, 1986; Allain *et al.*, 1986; Gaines *et al.*, 1987), although this normally subsides on development of specific antibodies. However, antigen may subsequently reappear, often in association with disease progression (Allain *et al.*, 1986). One problem with quantifying circulating HIV by this method is that the viral proteins may be partially or completely complexed with anti-HIV antibody, and thus be undetectable by the conventional antigen assay, but positive after prior dissociation of the immune complexes (Lange *et al.*, 1987; Ujhelyi *et al.*, 1987). Antigen levels therefore reflect the balance of virus and antibody production. In this study, high levels of viral RNA sequences in antigen-negative plasma samples were frequently found (patients 77, 74, 95, 70, and 72), yet similar or lower amounts of circulating RNA have been found in other plasma samples that are antigen-positive (patients 56, 28, 79, 82, and 87). This is a clear indication that differences in antibody levels do play a part in the

variability in detection of p24 antigen. In other studies, a similar lack of correlation between p24 antigen levels and titres of infectious virus in plasma has also been reported (Ho *et al.*, 1989; Coombs *et al.*, 1989).

Compared with the p24 antigen assay, the RNA PCR method provides a direct way to detect and quantify virus production regardless of immune complex formation, hence it may provide a better marker for the progression of disease. On average, HIV RNA was more abundant in the plasma of patients with more advanced disease compared with asymptomatic (Table 3.2). However, a wide range in the amount of cell-free HIV RNA was found among patients in similar stages of disease (between 200 to 8.5 x 10^3 in CDC stage II patients studied, see Table 3.2). These results can be explained by a variation in the level of p24 antibody as concluded following a recent study of antigen levels after dissolution of immune complexes. Total levels of p24 antigen, both free and immune complexed, have been found to vary little during the course of primary infection and subsequently (Goudsmit *et al.*, 1986; Allain *et al.*, 1986; Gaines *et al.*, 1987). Therefore, the relative high levels of HIV RNA, and/or reappearance of p24 antigenaemia at the later stage of an HIV infection may be largely a consequence of a reduction in levels of specific antibody (Nishanian *et al.*, 1990).

The lack of association between the levels of HIV-1 in plasma and the CD4+ cell counts is really unexpected. This result is consistent with some investigators' results (Ragni *et al.*, 1989; Clumeck *et al.*, 1989) but different from others' where the positive correlation has been found between the increasing amount of plasma levels of HIV-1 and the depletion of CD4+ cells (Goedert *et al.*, 1987; Laga *et al.*, 1989).

There maybe several explanations for the lack of correlation between the plasma viral load and CD4+ cell levels. However, one that is most likely may be that the quantitative feature of HIV-1 itself will not be the sole responsible factor for the CD4+ cell depletion. The differences in viral phenotypes, such as viral tropism, cytopathogenicity and virulence, have been noted for some time (Cheng-Mayer *et al.*, 1988; Tersmette *et al.*, 1989*a, b*). Viruses with different phenotypic features will therefore have different impact on the ultimate fate of infected CD4+ cells; less cytopathic variants will permit relative longer life-span of infected CD4+ cell whereas more virulent variants may kill infected CD4+ cell in a very short period of time. Thus, the quantitative as well as qualitative features of HIV-1 will therefore jointly determine the fate of infected CD4+ cells. As one of the consequences, the increasing amount of HIV-1 in plasma is not necessarily reflected by the substantial drop of CD4+ cells in the blood stream.

The levels of HIV in patients' plasma, whether they were from CDC group II or from patients undergoing antiviral treatment, were much higher than previously estimated (Table 3.2). Zidovudine treatment has previously been shown to decrease the amount of cell-free circulating HIV initially both in plasma of infected humans (Ho *et al.*, 1989) and of severe combined immunodeficiency infected mice (McCune *et al.*, 1990). The high concentration of cell-free circulating HIV in our patients, who have been undergoing zidovudine therapy for over 10 months, may imply that some resistant viral strains have emerged.

The most striking feature of the results is the high level of cell-free HIV found in some CDC group II patients (p56, 70 and 82, see Table 3.2). Serial samples

from one CDC group II patient (p82) without antiviral treatment showed persistently high levels of plasma virus for several years and rapid turnover of sequence variants (see chapter 4). The detection of high levels of cell-free HIV from the plasma of both CDC group II and IV patients suggests that viral replication occurs continuously throughout the course of an HIV infection. There is no virological evidence for a 'latent' period.

3.4.2 Detection of HIV RNA from Factor VIII Concentrates This is the first direct demonstration of contamination of factor VIII by HIV-1. Two out of eight batches of factor VIII concentrate were positive for HIV-1 RNA by PCR; in one case with *env* primers and the other with *pol* primers. Both were confirmed by sequencing the PCR product. The amount of HIV RNA present in factor VIII is very low (2.5 copies per ml), and close to the threshold of detection. These two factor VIII sequences, which were both found in commercial products, are distinct from those of any published HIV isolates, but are more closely related to North American variants than to African ones. One problem with detection of HIV-1 RNA in blood products, such as those factor VIII or IX or immunoglobulin, is the extremely low levels of viral RNA present in these blood products. If viral RNA could be readily detected and subsequently sequenced from these blood products, it would surely provide clearer picture of virus population in the original contaminated source, and therefore will help us to understand the viral sequential changes which enable viruses to persist in the face of potent anti-HIV immune response.

3.4.3 Detection and Quantitation of HCV RNA from Factor VIII Concentrates

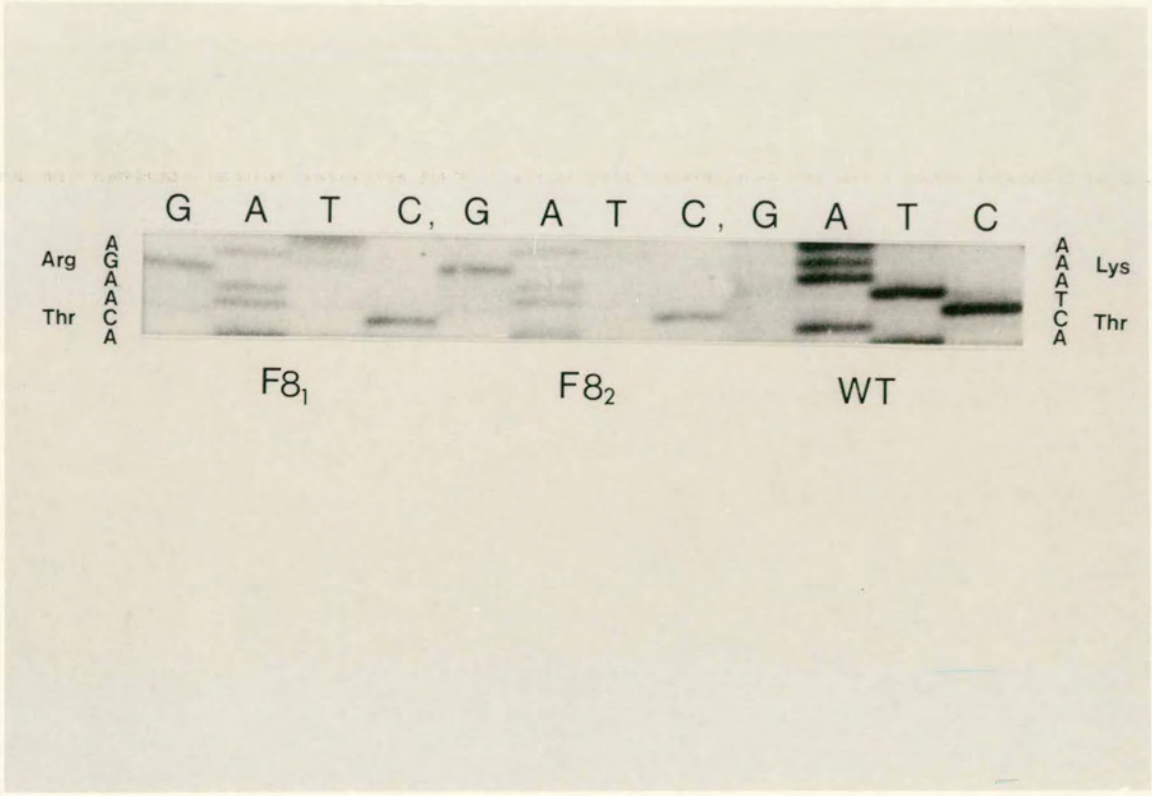
Previous studies by ourselves and others have detected hepatitis C virus (HCV) RNA sequences in factor VIII using similar methods (Simmonds *et al.*, 1990*b*; Garson *et al.*, 1990). In fact, out of the eight batches of factor VIII concentrates tested for HIV-1 RNA in this study, all six of the commercially-derived batches contain detectable amounts of HCV RNA, including the two positives for HIV RNA (the two National Health Service-derived batches were negative for HCV RNA). The amounts of HCV RNA were higher than HIV; in two batches that were quantified by limiting dilution, between 30,000 and 100,000 copies of HCV RNA per ml were found (Simmonds *et al.*, 1990*b*). There are many possible explanations for the difference in the concentration of HIV and HCV RNA presented. In one of our studies (Dr. Henry Watson, personal communication), comparison of the levels of HIV with those of HCV in plasma have shown that HCV is present in 10 and 100-fold greater amounts in infected individuals. Second, the prevalence of HCV infection may be higher in paid donors. Third, HIV may be less stable during the factor VIII fractionation process than HCV, or may be excluded with greater efficiency.

Higher titres of HCV RNA present in factor VIII concentrates may partially explain the higher rate of HCV infection in haemophiliac patients (Simmonds *et al.*, 1990*b*). Our cross-section study has shown that the prevalence of HCV infection is much higher in haemophiliacs infected with HIV compared to those not infected (Dr. Henry Watson, personal communication). Haemophiliacs who were HCV PCR-positive and HCV antibody-negative all had AIDS and low CD4 cell counts (<200 x 10^9/l) (Dr. Henry Watson, personal communication). HCV negative antibody

results may be due to either the immunodeficient state caused by infection of HIV or to high rates of HCV viral protein expression that could adsorb circulating antibody by immune complex formation.

3.4.4 Proposed Substitution for AZT Resistance Was a Pre-existing Polymorphism Larder *et al.*, firstly reported that isolates of HIV, taken from patients who had been taking AZT for more than 6 months, were resistant to AZT *in vitro* (Larder *et al.*, 1989*a, b, c*). Analysis of the coding region of reverse transcriptase (RT) from clinical isolates by nucleotide sequencing led to the discovery that multiple common nucleotide changes were associated with this resistance (Larder *et al.*, 1989*b*). These changes conferred specific amino acid substitutions in RT at the positions 67, 70, 215 and 219 (for more details see section 1.2.4). It was noted from the *pol* sequences obtained from the factor VIII concentrates batch no.8 that one of the four proposed AZT resistant mutations was found at position 70 in the RT domain (Figures 3.5 and 3.6). The discovery of this substitution in factor VIII concentrates prepared before the use of zidovudine as an antiviral agent suggests that proposed mutation at position 70 is a pre-existing polymorphism. Recently, AZT-resistant virus was detected in PBMCs of a few patients who had never received AZT, and therefore confirm our observation of pre-existing polymorphism at proposed sites which conferring AZT-resistance (Mohri *et al.*, 1993).

Figure 3.6 Polyacrylamide sequencing gel showing the amino acid replacement from Lysine to Arginine at position 70 in HIV pol gene, which is proposed to be associated with resistance to zidovudine, is a pre-existing polymorphism. $F8_1$ and $F8_2$ represent two molecules from factor VIII concentrate No.8. WT is a wildtype sequence.


G A T C, G A T C, G A T C
Arg
Thr
AGAACA
AAATCA
Lys
Thr
F8$_1$
F8$_2$
WT

CHAPTER 4

SEQUENTIAL SEQUENCE VARIATION OF HIV-1 *ENV* GENE IN PLASMA VIRAL POPULATION OF PATIENT 82

4.4 DISCUSSION

4.4.1 Effective Immune Response at Early Stage of HIV Infection

4.4.2 *In Vivo* Sequence Evolution of the V3 and V4 regions in Patient 82

4.4.3 Constraints on the V3 Sequences

4.4.4 Long-term Persistence of Seroconversion Sequences

4.4.5 Origin of Plasma Viral Population

4.1 Summary

In an investigation of the evolution of the envelope gene of HIV-1 during the course of an infection, a total of 89 V3 and 114 V4 viral RNA sequences were obtained from serial samples from an HIV-1 positive haemophilic. Sequence analysis of these two regions reveals that there are complex evolutionary patterns. For the V3 region, sequence variation has been found in both the V3 loop and the flanking regions. However, the frequent changes were mostly accumulated at sites in the V3 loop which have been proposed to be the targets recognised by the immune system. Moreover, selective constraints on the V3 region, particularly in the V3 loop, have also been observed, indicated by the dramatic fluctuations in the samples sizes of sequence variants, and extensive convergent evolution has been detected (Holmes, *et al.*, 1992). These findings strongly suggest that although there is selection for replacement of amino acids which may alter the B and T cells epitopes to evade immune recognition, there is also a severe selective constraint as to which amino acids are functionally viable within these regions.

Phylogenetic analysis of the V3 region reveals that there are several different evolutionary lineages in the plasma population after three years of infection. However, only two of them reach high frequency in the subsequent years. The variants found in the later stage of infection can all be identified as the progeny of the early sequence variants found at seroconversion.

Sequence variation in the V4 region is slightly different. Apart from frequent amino acid replacement through time, length variation is also observed in this region. The length variation of PCR amplified cDNA and DNA in V4 region was

investigated and a distinctive length pattern for coexisting variants in each sample was identified. In V4, each succeeding sequence type is not obviously more related to those that come before or after it than they are to the sequences obtained at the early stage of infection. This finding suggests that the evolution of viral RNA V4 sequences in the plasma population of this patient is different from those of the V3 region, and that the plasma viraemia may be contributed by virus variants in the solid tissues such as lymph node, brain, spleen, etc. In addition, there are significant differences between the frequencies of sequence variants in proviral DNA and viral RNA populations from the same sample, indicating that at any one time point, the predominant plasma virus variants may be antigenically distinct from some of those viruses encoded by HIV DNA sequences in PBMCs.

4.2 Introduction

The surface glycoprotein of HIV-1, gp120, like those of other retroviruses, shows considerable sequence diversity between independent isolates. Independent isolates from North American differ in almost 10-20% of amino acid residues in the *env* gene, whereas isolates from Africa can differ in up to 25-30% of these sites (Balfe *et al.*, 1990). In common with visna virus and equine infectious anaemia virus (EIAV) (Clements *et al.*, 1988), HIV also exhibits variability within infected individuals. Sequence analysis of consecutive isolates obtained by *in vitro* culturing with primary lymphocytes from the same individuals showed rapid sequence change over time (Meyerhans *et al.*, 1989; Simmonds *et al.*, 1990*a* and 1991). Genetic divergence has also been observed even within the same isolate, suggesting that the HIV isolate is a

heterogenous population (Putney *et al.*, 1990*a, b*).

Recently, it has been demonstrated that some parts of the *Env* region are crucial for viral functions. The major CD4 binding domain, through which the virus facilitates its attachment to the membrane of target cells, has been located between the V4 and V5 hypervariable regions (Lasky *et al.*, 1987). A number of reports have recently suggested that major determinants for macrophage and T-cell tropism are also located in a region of gp120 which encompasses the third hypervariable region V3 (Cheng-Mayer *et al.*, 1990*a, b*; Hwang *et al.*, 1991; Westervelt *et al.*, 1991 and 1992). The V3 region may also play an assisting role in membrane fusion after gp120 binding to CD4 (Freed *et al.*, 1991 and 1992).

A highly antigenic epitope for both B and T cells has been identified in the third hypervariable region (V3) (Rusche *et al.*, 1988; Palker *et al.*, 1988; Javaherian *et al.*, 1989). A peptide encompassing the V3 domain elicits type-specific neutralizing antibody titers that are comparable to those elicited by the entire gp160 or gp120, and most neutralizing antibodies elicited by immunization with recombinant gp160 or gp120 proteins can be absorbed with a V3 peptide (Javaherian *et al.*, 1989). This segment of the envelope has therefore been termed the principal neutralization determinant (PND) of the envelope protein. The coincidence of the hypervariable nature of the V3 region and its high antigenicity may suggest that the high rate of changes in the V3 region could generate a succession of mutants which allow the variants to escape from of immune surveillance. This notion was recently supported by experiments showing an *in vivo* emergence of virus variants that were resistant to neutralization by autologous sera (Albert *et al.*, 1990; Montefiori *et al.*, 1991).

Moreover, an HIV-1 'escape' mutant has also been selected by neutralizing antibody *in vitro*, due to selection for a point mutation in the V3 loop (McKeating *et al.*, 1989).

In the early studies of sequence variation, viral sequences were largely obtained from isolates cultured with primary lymphocytes (Hahn *et al.*, 1986; Starcich *et al.*, 1986). However, it has been shown that the *in vitro* propagation of viral isolates can reduce genetic diversity in the viral envelope region (Meyerhans *et al.*, 1989; Kusumi *et al.*, 1992). Therefore, studies of genetic variation based on the sequences obtained from *in vitro* cultured isolates may not truly reflect sequence variation *in vivo*.

In the current work, the sequential variation in HIV-1 *env* genes in plasma viral population has been studied. Using the highly sensitive and reliable RNA PCR method developed previously in this work (see chapter 3), viral RNA was detected directly from patient plasma samples, amplified and sequenced without further *in vitro* manipulation. By this means, not only can errors introduced by *Taq* polymerase during the amplification process be avoided, but also any bias due to *in vitro* culturing (Meyerhans *et al.*, 1989; Simmonds *et al.*, 1990*a*). As such, the sequences obtained by this direct detection and sequencing technique will be more representative of those *in vivo*. Starting from seroconversion at year 0 (1984), a total of six serial plasma samples were collected from a single HIV-1 positive haemophiliac at years 3 (1987), 4(1988), 5(1989), 6(1990) and 7(1991). Sequence analysis has largely concentrated on the V3 and V4 hypervariable regions of the envelope gene.

4.3 Results

The patient studied was a haemophiliac(p82) who was infected with HIV-1 from a contaminated factor VIII concentrate and seroconverted in 1984 (Ludlam *et al.*, 1985). This patient was seroconverted around March 1984, and asymptomatic at the time of study and has never undergone any antiviral therapy. The virological and immunological data on this patient are shown in Table 4.1. The levels of plasma viraemia were quantified by RNA PCR method previously described in this work. The other data in the Table 4.1 are kindly provided by Dr. Peter Simmonds [proviral DNA load]; Mrs. Selma Rebus [p24 antigen level] (Department of Medical Microbiology, University of Edinburgh) and Dr. Henry Watson [CD4+ count] (Department of Haematology, Royal Infirmary of Edinburgh). Nucleotide sequences were aligned using LINEUP and PILEUP programs on the University of Wisconsin GCG package (Devereux *et al.*, 1984). The nucleotide distance matrices were estimated by program DNADIST. Maximum likelihood and neighbor-joining trees were then inferred using the DNAML and Neighbor programs respectively from the PHYLIP (version 3.4) of Felsenstein (see section 2.2.12).

4.3.1 Rapid Changes of Plasma Viraemia During the Early Stage of HIV Infection

Quantitative analysis of plasma viraemia was carried out using previously developed RNA based PCR method (see Chapter 3 for more detail). At seroconversion, the levels of plasma viraemia were exceptionally high at approximately 10^8 viral RNAs per ml of plasma (see Table 4.1). However, the subsequent samples all showed a substantially lower viral load in the plasma, nearly 4 orders of magnitude less than

Table 4.1 Standard virological and immunological markers of HIV infection in p82. Viral load in plasma and PBMCs was quantified by limiting dilution of viral cDNA and DNA prior to double PCR amplification. Proviral DNA load and p24 antigen level in plasma were kindly provided by Dr. P. Simmonds and Mrs. Selma Rebus respectively (Department of Medical Microbiology, University of Edinburgh). CD4+ count is kindly provided by Dr. Henry Watson (Department of Haematology, Edinburgh Royal Infirmary). [a] Time from first positive serum sample. [b] Detection of serum antigen by ELISA (>15 pg/ml; Dupont). NA, not applicable [c] Proportion of PBMCs bearing provirus. ND, not done.

Sample	Times (months)[a]	$CD4^+$ Lymphocytes (x 10^9/l)	p24 antigen[b]	Provirus-bearing PBMCs[c]	Viral RNA in plasma
1983	-14	1.45	NA	NA	NA
1984	0	0.93	+	ND	1.0×10^8
1987	36	0.53	-	1/2000	2.8×10^4
1988	43	0.34	-	1/2270	5.7×10^3
1989	56	0.65	+	1/700	8.5×10^3
1990	70	0.10	-	ND	8.2×10^3
1991	79	0.10	-	ND	5.8×10^3

that at seroconversion. This drop in viral titre implies an effective immune response during the early stage of HIV infection. However, a persistently high level of plasma viraemia during the asymptomatic stage was also observed ranging from 5.7-8.5 x 10^3 viral RNAs per ml of plasma (see Table 4.1). This suggests that despite the strong immune reaction, viral replication continues. There is no virological silent phase corresponding to the latent period observed clinically.

4.3.2 Nucleotide Sequences from Plasma Viral RNAs A total of 89 nucleotide sequences of the V3 region and 114 of the V4 region were obtained directly from double PCR amplified single cDNA molecules (Figure 4.1 [a] and [b], respectively). Each sequence represents an individual viral genome (or RNA molecule), isolated by limit dilution of cDNAs prior to PCR amplification (see section 2.2.7). In many cases, the V3 and V4 sequences were obtained from the same genome (indicated by the identical sequence name in the figures) and can therefore be used for studies of linkage and recombination between these two regions. The sequences were aligned and gaps (indicated by '-') were introduced where it was necessary to preserve the alignment.

4.3.3 Low Frequency of Inactivating Substitutions As shown in Figure 4.1 the sequences in the V3 and V4 regions of *env* gene of patient 82 exhibit an enormous degree of variation. Frequent nucleotide substitutions have been observed in both hypervariable regions. In addition deletion or insertion events have been identified, particularly in the V4 hypervariable region (see Figure 4.1 [b]). In many cases, a

Figure 4.1. Viral RNA sequences from patient 82. Sequences from (a) the V3 region, and (b) from the V4 region. Each sequence represents a single viral RNA sequence obtained directly from plasma. Nucleotide positions are numbered according to HIV_{HXB2}. Only nucleotides that differ from those obtained at year 0 are shown. Gaps were introduced to preserve the alignment and indicated by dashes. Dots denote to the identical nucleotides shown above. The same name was given to the V3 and V4 sequences when they were obtained from the same molecule. Year 0 = 1984; year 3 = 1987; year 4 = 1988; year 5 = 1989; year 6 = 1990; year 7 = 1991.

(a)

Year	Sequences	7051	7110
	Z82r001	TTTCACGGAC AATGCTAAAA CCATAATAGT ACAGCTGAAG GAATCTGTAG AAATTAATTG	
	Z82r00a	....	
	Z82r00b	....	
	Z82r00c	....	
	Z82r00d	....	
0	Z82r00e	....	
	Z82r00f	....	
	Z82r00g	....	
	Z82r00h	....	
	Z82r00i	....	
	Z82r00j	....	
	Z82r00k	....	
	Z82r003	......AA.. ....T	
	Z82r004	......AA.. ....T	
	Z82r005	....G.A... ....T	
	Z82r007	....G.A... ....T .G........	
	Z82r008	......A...	
	Z82r009	......A... .... T.........	
	Z82r01	......A...	
3	Z82r02	......A...	
	Z82r03	......A... .... T.........	
	Z82r04	......A...	
	Z82r05	......A... ....T .... T.........	
	Z82r06	......A... .... T.........	
	Z82r07	......A... ....T .... T.........	
	Z82r08	......A... ....T .... T.........	
	Z82r09	......A... .... T.........	
	Z82r10	......A... ....T .... T.........	
	Z82r11	......A... ....T .... T.........	
	Z82r12	......A... ....T .... T.........	
	Z82r13	......A... ....T .... T.........	
	Z82r15	......A..T ....T .... T.........	
4	Z82r16	......A... ....T .... T.........	
	Z82r17	......A... ....T .... T.....C...	
	Z82r18	......A..T ....T .... T.........	
	Z82r19	......A... ....T .G........ T.........	
	Z82r20	......A... ....T .... T.........	
	Z82r21	......A... ....T .... T.........	
	Z82r22	......A... ....T .... T.........	
	Z82r24	......A... ....T .....C.... T.....C...	
	Z82r25	......A... ....T .... T.........	
	Z82r26	......A... ....T .... T.....C...	
	Z82r27	......A... ....T .... T.........	
	Z82r28	......A... ....T .... T.........	
	Z82r29	....G.A... ....T ...C...... T.........	
	Z82r30	....G.A... ....T ......AC.. T.........	
	Z82r31	....G.A... ....T ......AC.. T.........	
	Z82r32	......A... ....T .... T.........	
5	Z82r33	...A..A... ....T .... T.........	
	Z82r34	...A..A... ....T .... T.........	
	Z82r35	......A... ....T ......AC.. T.........	
	Z82r37	.C....A... ....T .... T.........	
	Z82r38	......A... ....T .... T.........	
	Z82r39	...T..A... ....T .... T.........	
	Z82r40	......A... ....T .... T.........	
	Z82r41	......A... ....T .... T.........	
	Z82r42	......A... ....T .... T.....C...	
	Z82r43	...T..A... ....T .... T.........	
	Z82r44	......A... ....T .... T.........	
	Z82r45	...T..A... ....T .... T.........	
	Z82r46	...T..A... ....T .... T.........	
	Z82r70	......A..T ....T .... T.........	
	Z82r71	......A... ....T .... T.........	
	Z82r72	......A... ....T .... T.........	
	Z82r75	......A... ....T .... T.........	
	Z82r76	......A... ....T .... T.........	
	Z82r77	......A... ...A...... ....T .... T.........	
	Z82r78	......A... ...A...... ....T .... T.........	
6	Z82r79	......A..T ...A...... ....T .... T.....C...	
	Z82r80	......A... ....T .... T.........	
	Z82r81	......A... ....T .C........ T.....C...	
	Z82r82	......A... ....T .... T.........	
	Z82r83	......A... ...A...... ....T .... T.....C...	
	Z82r84	......A... ....T .C........ T.....C...	
	Z82r85	......A... ....T .... T.........	
	Z82r86	......A... ....T .... T.........	
	Z82r91	......A... ...A...... ....T A........ T.....C...	
	Z82r92	......A... ....T .... T.........	
	Z82r93	......A... ....T .... T.........	
	Z82r94	....G. ....T .... T.........	
	Z82r95	......A... ....T .... T.........	
	Z82r96	......A... ....T .... T.........	
7	Z82r97	......A... ....T .... T.........	
	Z82r98	......A... .... .T........ ..T......T A........ T.........	
	Z82r99	......A... ....T .... T.........	
	Z82r100	......A... ...A...... ....T .... T.........	
	Z82r101	......A... ....T .... T.........	
	Z82r102	......A... ....T .... T.........	
	Z82r103	......A... ....T .... T.........	

Year	Sequences	7111					7169
	Z82r001	TACAAGACCC	AACAACAATA	CAAGAAAAAG	TATACATATA	GGACCAGGAA	GAGCATTTTA
	Z82r00a	....	....	....	....	....	....
	Z82r00b	....	....	....	....	....	....
	Z82r00c	....	....	....	....	....	....
	Z82r00d	....	....	....	....	....	....
0	Z82r00e	....	....	....	....	....	....
	Z82r00f	....	....	....	....	....	....
	Z82r00g	....	....	....	....	....	....
	Z82r00h	....	....	....	....	....	....
	Z82r00i	....	....	....	....	....	....
	Z82r00j	....	....	....	....	....	....
	Z82r00k	....	....	....	....	....	....
	Z82r003	....	....	....	.....C....	........G.	....
	Z82r004	....	....	....	....	........GG	....
	Z82r005	....	....	....	....	........GG	....
	Z82r007	....	....	....	....	........GG	....
	Z82r008	....	....	....	....	........G.	....
	Z82r009	....	....	....	.....C....	........G.	....
	Z82r01	....	....	....	.....C....	........G.	....
3	Z82r02	....	..T.......	....	.....C....	........G.	....
	Z82r03	....	....	....	....	........GG	....
	Z82r04	....	....	....	....	........G.	....
	Z82r05	....	....	........G.	....	........G.	.T........
	Z82r06	....	....	....	.....C....	........G.	....
	Z82r07	....	....	....	....	........G.	....
	Z82r08	....	....	....	....	........G.	....
	Z82r09	....	....	....	.....C....	........G.	....
	Z82r10	....	....	....	....	........G.	.....G....
	Z82r11	....	....	....	....	........G.	....
	Z82r12	....	....	........G.	....	........G.	.T........
	Z82r13	....	....	........G.	....	........G.	.....G....
	Z82r15	....	....	........G.	....	........G.	.T........
4	Z82r16	....	....	....	G.........	........G.	.....G....
	Z82r17	....	....	....	G.........	........G.	.....G....
	Z82r18	....	....	........G.	....	....	.T........
	Z82r19	....	....	....	G.........	........G.	.....G....
	Z82r20	....	....	....	G.........	........G.	.....G....
	Z82r21	....	....	....	G.........	........G.	.....G....
	Z82r22	....	....	........G.	....	........G.	.T........
	Z82r24	....	....	....	G...T.....	........G.	.....G....
	Z82r25	....	....	........G.	....	........G.	.T...G....
	Z82r26	....	....	....	G...T.....	........G.	.T...G....
	Z82r27	....	....	........G.	....	........G.	.T........
	Z82r28	....	....	........G.	....	........G.	.T........
	Z82r29	....	....	........G.	....	....	.T........
	Z82r30	....	....	........G.	....	........G.	.T........
	Z82r31	....	....	........G.	....	........G.	.T........
	Z82r32	....	....	........G.	....	........G.	.T........
5	Z82r33	....	....	........G.	....	........G.	.T........
	Z82r34	....	....	........G.	....	........G.	.T........
	Z82r35	....	....	........G.	....	........G.	.T........
	Z82r37	....	....	........G.	....	........G.	.T........
	Z82r38	....	....	........G.	....	........G.	.T........
	Z82r39	....	....	........G.	....	........G.	.T........
	Z82r40	....	....	........G.	....	........G.	.T........
	Z82r41	....	....	........G.	....	........G.	.T........
	Z82r42	....	....	....	G...T.....	........G.	.....G....
	Z82r43	....	....	........G.	....	........G.	.T........
	Z82r44	....	....	........G.	....	........G.	.T........
	Z82r45	....	....	........G.	....	........G.	.T........
	Z82r46	....	....	........G.	....	........G.	.T........
	Z82r70	....	......T...	....	G..TGG....	........G.	..T..G....
	Z82r71	....	....	........G.	....	........G.	.T........
	Z82r72	....	....	........G.	....	........G.	.T...G....
	Z82r75	....	......T...	....	G..TGG....	........G.	..T..G....
	Z82r76	....	....	........G.	....	........G.	....
	Z82r77	....	......T...	....	G..TGG....	........G.	..T..G....
	Z82r78	....	....	........G.	....	........G.	.....G....
6	Z82r79	....	....	....	G...T.....	....	.....G....
	Z82r80	....	....	........G.	....	........G.	.T........
	Z82r81	....	....	....	G...T.....	........G.	.....G....
	Z82r82	....	....	........G.	....	........G.	.T........
	Z82r83	....	....	....	G...T.....	....	.....G....
	Z82r84	....	....	....	G...T.....	........G.	.....G....
	Z82r85	....	....	........G.	....	........G.	.T........
	Z82r86	....	....	........G.	....	........G.	.T........
	Z82r91	....	....	....	G...T.....	........G.	.....G....
	Z82r92	....	....	........G.	....	........G.	.T........
	Z82r93	....	....	........G.	....	........G.	.T........
	Z82r94	....	..T.......	........G.	....	........G.	.T........
	Z82r95	....	....	........G.	....	........G.	.T........
	Z82r96	....	....	........G.	....	........G.	.T........
7	Z82r97	....	....	........G.	....	........G.	.T........
	Z82r98	....	....	........G.	....	........G.	.T........
	Z82r99	....	....	........G.	....	....	.T........
	Z82r100	....	......T...	....	G..TGG....	........G.	..T..G....
	Z82r101	....	..T.......	........G.	....	........G.	.T........
	Z82r102	....	....	....	....	........G.	....
	Z82r103	....	....	........G.	....	....	.T........

Year	Sequences	7170					7229
	Z82r001	TACAACAGGA	GAAATAATAG	GAGATATAAG	ACAAGCACAT	TGTAACCTTA	GTAGAGCAAA
	Z82r00a	....	....	....	....	....	....
	Z82r00b	....	....	....	....	....	....
	Z82r00c	....	....	....	....	....	....
	Z82r00d	....	....	....	....	....	....
0	Z82r00e	....	....	....	....	....	....
	Z82r00f	....	....	....	....	....	....
	Z82r00g	....	....	....	....	....	....
	Z82r00h	....	....	....	....	....	....
	Z82r00i	....	....	....	....	....	....
	Z82r00j	....	....	....	....	....	....
	Z82r00k	....	....	....	....	....	....
	Z82r003	....	C.........	....	....	....	....
	Z82r004	....	C.........	....	....	....	....
	Z82r005	....	C.........	....	....	......A...	....
	Z82r007	....	C.........	....	....	....	....
	Z82r008	....	..C....C..	....	....	....	....
	Z82r009	....	..C....C..	....	....	....	....
	Z82r01	....	..C.......	....	....	....	....
3	Z82r02	....	C.........	....	....	....	....
	Z82r03	....	C.........	....	....	....	....
	Z82r04	....	..C.......	....	....	......A...	....
	Z82r05	.G........	..C.......	....	....	....	....
	Z82r06	....	C.........	....	....	....	....
	Z82r07	....	.G........	....	....	......A...	....
	Z82r08	....	.G........	....	....	......A...	....
	Z82r09	....	C.........	....	....	....	....
	Z82r10	....	C.........	....	....	....	....
	Z82r11	C.........	.G........	....	....	......A...	....
	Z82r12	.G........	.G........	....	....	......A...	....
	Z82r13	........A.	C.........	..A.......	....	......A...	....
	Z82r15	.G........	.G........	....	G.........	......A...	....
4	Z82r16	........A.	C.........	..A.......	....	......A...	....
	Z82r17	........A.	C.........	..A.......	....	......A...	....
	Z82r18	.G........	CGG.......	....	....	......A...	....
	Z82r19	........A.	C.........	..A.......	....	......A...	....
	Z82r20	........A.	C.........	..A.......	....	....	....
	Z82r21	........A.	C.........	..A.......	....	....	....
	Z82r22	.G........	.G........	....	G.........	......A...	....
	Z82r24	........A.	C.........	..A.......	....	......A...	....
	Z82r25	.G........	.G........	....	....	......A...	....
	Z82r26	........A.	C.........	..A.......	....	......A...	....
	Z82r27	.G........	.G........	..A.......	....	......A...	....
	Z82r28	.G........	..C.......	....	....	......A...	....
	Z82r29	.G........	.G........	....	....	......A...	....
	Z82r30	.G........	.G........	....	....	......A...	....
	Z82r31	.G........	.G........	....	....	......A...	....
	Z82r32	.G........	.G........	....	....	......A...	....
5	Z82r33	.G........	.G........	....	....	......A...	....
	Z82r34	.G........	.G........	....	....	......A...	....
	Z82r35	.G........	.G........	....	....	......A...	....
	Z82r37	.G........	.G........	....	G.........	......A...	....
	Z82r38	.G........	.G........	....	....	......A...	....
	Z82r39	.G........	.G........	....	....	......A...	....
	Z82r40	.G........	.G........	....	....	......A...	....
	Z82r41	.G........	.G........	....	....	......A...	....
	Z82r42	........A.	C.........	..A.......	....	......A...	....
	Z82r43	.G........	.G........	....	....	......A...	....
	Z82r44	.G........	.G........	..A.......	....	......A...	....
	Z82r45	.G........	.G........	....	....	......A...	....
	Z82r46	.G........	.G........	....	....	......A...	....
	Z82r70	....G...A.	C.........	..A.......	....	......A...	....
	Z82r71	.G........	..C.......	....	....	.....T....	....
	Z82r72	.G........	..C.......	....	G.........	......A...	....
	Z82r75	....G...A.	C.........	..A.......	....	......A...	....
	Z82r76	.G........	..C.......	....	G.........	......A...	....
	Z82r77	....G...A.	C.........	..A.......	....	......A...	....
	Z82r78	....	..C.......	....	....	......A...	....
6	Z82r79	........AT	C.........	....	....	......A...	....
	Z82r80	.G........	..C.......	....	....	......A...	....
	Z82r81	........AT	C.........	..A.......	....	......A...	....
	Z82r82	.G........	..C.......	....	G.........	......A...	....
	Z82r83	........AT	C.........	..A.......	....	......A...	C
	Z82r84	........AT	C.........	..A.......	....	......A...	C
	Z82r85	.G........	..C.......	....	G.........	......A...	....
	Z82r86	.G........	..C.......	....	G.........	......A...	....
	Z82r91	........AT	C.........	..A.......	....	......A...	....
	Z82r92	.G........	..C.......	....	....	......A...	....
	Z82r93	.G........	..C.......	....	G.........	......A...	....
	Z82r94	.G........	..C.......	....	G.........	......A...	....
	Z82r95	.G........	..C.......	....	....	......A...	....
	Z82r96	.G........	..C.......	....	G.........	......A...	....
7	Z82r97	.G........	..C.......	....	....	......A...	....
	Z82r98	.G........	..C.......	....	G.........	......A...	....
	Z82r99	.G........	..C.......	....	....	......A...	....
	Z82r100	....G...A.	C.........	..A.......	....	......A...	.......C..
	Z82r101	.G........	..C.......	....	....	......A...	....
	Z82r102	C.........	.G........	....	....	......A...	....
	Z82r103	.G........	..C.......	....	....	......A...	....

Year	Sequences	7230	7283
	Z82r001	ATGGACTGAC ACTTTAAGAC AGATAGTTAT GAAATTAAGA GAACAATTTG GGAA	
	Z82r00a	....	
	Z82r00b	....	
	Z82r00c	....	
	Z82r00d	....	
0	Z82r00e	....	
	Z82r00f	....	
	Z82r00g	....	
	Z82r00h	....	
	Z82r00i	....	
	Z82r00j	....	
	Z82r00k	....	
	Z82r003	.....A...A .... A......G..	
	Z82r004	.....A...A .... A......... .... A...	
	Z82r005	.....A...A ....A.. .... A......... .... A...	
	Z82r007	.....A...A ....A.. .... A......G.. A.........	
	Z82r008	.....A...A ....A.. .... A......G..	
	Z82r009	.....A...A ....A.. .... A......G..	
	Z82r01	.....A...A .... A......G.. A.........	
3	Z82r02	.....A...A .... A......G..	
	Z82r03	.....A...A .... A......G..	
	Z82r04	.....A.... ......G..A ....G..	
	Z82r05	.....A...A ....A.. .... A......G.. ....A	
	Z82r06	.....A...A .... A......G..	
	Z82r07	.....A...A ....A.. .... A......... .... A...	
	Z82r08	.....A...A ....A.. .... A......... .... A...	
	Z82r09	.....A...A .... A......G..	
	Z82r10	.....A...A .... A......G.. C	
	Z82r11	.....A...A ....A.. .... A.........	
	Z82r12	.....A...A ....A.. .... A......... .... A...	
	Z82r13	.....A...A ....A.. .... A......... .... A...	
	Z82r15	.....A...A ....A.. .... A......... ....A A...	
4	Z82r16	.....A...A ....A.. .... A......... .... A...	
	Z82r17	.....A...A ....A.. .... A......... .... A...	
	Z82r18	.....A...A ....A.. .... A......... .... A...	
	Z82r19	.....A...A ....A.. .... A......... .... A...	
	Z82r20	.....A...A .... A......G..	
	Z82r21	.....A...A .... A......G..	
	Z82r22	.....A...A ....A.. ....C A......... ....A A...	
	Z82r24	.....A...A ....A.. .... A......... .... A...	
	Z82r25	.....A...A ....A.. .... A.........	
	Z82r26	.....A...A ....A.. .... A......... .... A...	
	Z82r27	.....A...A ....A.. .... A.........	
	Z82r28	.....A...A ....A.. .... A......... .... A...	
	Z82r29	.....A...A ....A.. .... A.........	
	Z82r30	.....A...A ....A.. .... A.........	
	Z82r31	.....A...A ....A.. .... A.........	
	Z82r32	.....A...A ....A.. .... A......... ....A A...	
5	Z82r33	.....A...A ....A.. .... A.........	
	Z82r34	.....A...A ....AG. .... A.........	
	Z82r35	.....A...A ....A.. .... A......... ....A A...	
	Z82r37	.....A...A ....A.. ....C A......... ....A A...	
	Z82r38	.....A...A ....A.. .... A.........	
	Z82r39	.....A...A ....A.. .... A.........	
	Z82r40	.....A...A ....A.. .... A.........	
	Z82r41	.....A...A ....A.. ....C A.........	
	Z82r42	.....A...A ....A.. .... A......... .... A...	
	Z82r43	.....A...A ....A.. .... A.........	
	Z82r44	.....A...A ....A.. .... A.........	
	Z82r45	.....A...A ....A.. .... A.........	
	Z82r46	.....A...A ....A.. .... A.........	
	Z82r70	.....A...A ....A.. .... A.........	
	Z82r71	.....A...A ....A.. .... A.........	
	Z82r72	.....A...A ....A.. .... A......... .... .AG.	
	Z82r75	.....A...A ....A.. ....C. A......... .... A...	
	Z82r76	.....A...A ....A.. .... A.........	
	Z82r77	.....A...A ....A.. .... A......... .... AT..	
	Z82r78	.....A...A ....A.. ....C A.........	
6	Z82r79	.....A...A ....A.. ....C A......... ....A A...	
	Z82r80	.....A...A ......GA.. .... A......... ....G.	
	Z82r81	.....A...A ....A.. ....C A.........	
	Z82r82	.....A...A ....A.. ....C A.........	
	Z82r83	.....A.... ....A.. ....C A.........	
	Z82r84	.....A.... ....A.. .... A.........	
	Z82r85	.....A.A.A ......GA.. .... A.........	
	Z82r86	.....A.A.A ......GA.. ....C A.........	
	Z82r91	.....A...A ....A.. .... A......... .....G...A A...	
	Z82r92	.....A...A ......GA.A .... A.........	
	Z82r93	.....A.A.A ......GA.. .... A......... .... A...	
	Z82r94	.....A.A.A ......GA.. ....C A......... .... A...	
	Z82r95	.....A.A.A ......GA.. ....C A......... .... A...	
	Z82r96	.....A.A.A ......GA.. .... A.........	
7	Z82r97	.....A.A.A ......GA.. ....C A......... .... A...	
	Z82r98	.....A...A ....A.. .... A.........	
	Z82r99	.....A.A.A ......GA.. ....C A......... .... A...	
	Z82r100	.....A...A ....C. A......... ....A A...	
	Z82r101	.....A.A.A ......GA.. ....C A......... .... A...	
	Z82r102	.....A...A ....A.. .... A......... .... .A..	
	Z82r103	.....A.A.A ......GA.. ....C A......... .... .A..	

(b)

Year	Sequences	7376	7461
		TGGAATTCAA CA-------- ----CAACTT AATAGTACTT GGAATTCAAC ACAACTTAAT AGTGCTGGGA ATAATACTGA AGAAAA
	82r002	....AAT-- ---------- ---------- -----..... ......
	82r005	.... ---------- -----..... ......
	82r003	.... T......... ....T.. .... ---------- -----.G... ......
	82r004	....T.. ....C..- -----..... ......
0	82r00a	....T.. ....C..- -----..... ......
	82r00b	....T.. ....C..- -----..... ......
	82r00c	....T.. ....C..- -----..... ......
	82r00d	....T.. ....C..- -----..... ......
	82r00e	....T.. ....C..- -----..... ......
	82r5572	.......ATT AT..... .....---- ...G..... ....A.... ---------- -----..... ......
	82r5571	.......ATT .T..... .....---- ...G..... ....A.... ---------- -----....G ......
	82r5575	.......ATT .T..... .....---- ...G..... ...T..... .....A.... ---------- -----..... ......
	82r5577	.......ATT .T..... .....---- ...GA..... ...G..... .....A.... ---------- -----..... ......
	82r5579	.......ATT .T..... .....---- ...G...A.. ...G..... .....A.... ---------- -----..... ......
	82r5576	.......ATT .T..... .....---- ...G...A.. ...G..... .....A.... ---------- -----..... ......
	82r5573	.......ATT .T..... .....---- ...G...A.. ...G..... .....A.... ---------- -----..... ......
	82r5578	.......ATT .T..... .....---- ...G...A.. ...G..... .....A.... ---------- -----..... ......
3	82r5580	.......ATT .T..... .....---- ...G...A.. ...G..... .....A.... ---------- -----..... ......
	Z82r01	.......ATT TT..... .....---- ...G..... ....A.... ---------- -----....G ......
	Z82r02	.......ATT .T..... .....---- ...G...A.. ...G..... .....A.... ---------- -----..... ......
	Z82r03	.......ATT .T..... .....---- ...G...A.. ...C..... .....A.... ---------- -----....G ......
	Z82r04	.......ATT .T..... .....---- ...GA..... ...G..... .....A.... ---------- -----..... ....G.
	Z82r05	...G...ATT .T..... .....---- ....G. ....A..... ---------- -----..... ......
	Z82r06	.......ATT .T..... .....---- ...G..... ...G..... .....A.... ---------- -----..... ......
	Z82r07	.......ATT .T..... .....---- ...G..... ...G..... .....A.... ---------- -----..... ......
	Z82r08	.......ATT .T..... .....---- ...G..... ...T..... .....A.... ---------- -----..... ......
	Z82r09	.......ATT .T..... .....---- ...G...A.. ...G..... .....A.... ---------- -----..... ......
	82r122	......---- --TGGAATTA TTCT------ ...G...A.. ...G..... .....A.... ---------- -----....G ......
	82r1210	....G..ATT .T..... .....---- ...G...A.. ...G..... .....A.... ---------- -----....G ......
	82r125	....G..ATT .T..... .....---- ...G..... ....A.... ---------- -----....G ......
	82r1212	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	82r1213	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	82r1214	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	82r1223	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	82r126	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
4	Z82r10	....C..ATT .T..... .....---- ...GA..... ....A.... ---------- -----....G ......
	Z82r11	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r12	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r13	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r15	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r16	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r17	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.G ......
	Z82r18	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r19	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r20	....G..ATT .T..... .....---- ...G..... ....A.... ---------- -----....G ......
	Z82r21	....G..ATT .T..... .....---- ...G..... ....A.... ---------- -----....G ......
	82r821	....TGGGATTT AACA..... ....T.C A.....---- ---------- ---------- -----.GA.. ......
	82r825	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	82r8215	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	82r1091	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	82r1092	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	82r1094	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	82r1095	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	82r1097	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	82r1098	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	82r1096	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	82r1099	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	82r1093	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	82r1100	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	82r984	....TGGGATTT AACA..... ...G.....C A.....---- ---------- ---------- -----.AA.. ......
	82r9814	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	82r986	....TGGGATTT AACA...--- ---------C A.....---- ---------- ---------- -----.AA.. ......
	82r9810	....TGGGATTT AACA..... ....C A.G..---- ---------- ---------- -----.AAA. ...C..
	82r988	....TGGGATTT AACA..... ....C A.G..---- ---------- ---------- -----.AAA. ...C..
	82r9817	....TGGGATTT A...------ ....C A.....---- ---------- ---------- -----.AA.. ......
	82r9820	....TGGGATTT AACA....C. ....C A.....---- ---------- ---------- -----.AA.. ......
5	Z82r22	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r24	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r25	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r26	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r27	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r28	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r29	....TGGGATTT AACA.---.. ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r30	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r31	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r32	....TGGGATTT AACA..... ..CTACT..C A.....---- ---------- ---------- -----.AA.. ......
	Z82r33	....TGGGATTT AACA..... .....AG..C A.....---- ---------- ---------- -----.AA.. ......
	Z82r34	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r35	....TGGGATTT AACA..... GG.......C A.....---- ---------- ---------- -----.AA.. ......
	Z82r37	........G. ..TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r38	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r39	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r40	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r41	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r42	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r43	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r44	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r45	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......
	Z82r46	....TGGGATTT AACA..... ....C A.....---- ---------- ---------- -----.AA.. ......

6	Z82r70	....TGGGATTC AACA..... C.G.....- ---------- ---------- ---------- --------.. ....G.
	Z82r71	....C A.-------- ---------- ---------- -----.AA.. .....
	Z82r72	....TGGGATTT AACGA.T... G..G...A.C A.-------- ---------- ---------- -----GAA.G .....
	Z82r76	....GGGGATTT AACA..... .....C A.-------- ---------- ---------- -----.AA.. ..G...
	Z82r77	....TGGGATTT AACA..... .....C A.G..----- ---------- ---------- --------A. ..G...
	Z82r78	....TGGGATTT AACA..... ...G.....C A.G..----- ---------- ---------- -----.AA.G .....
	Z82r79	....TGGGATTT AACA..... ....A..A.A CTC.G----- ---------- ---------- -----.A.AG ....G.
	Z82r80	....A. ....C.---- ---------- ---------- ---------- --------.G .....
	Z82r81	....TGGGAGTT AACA..... .....C A.G..----- ---------- ---------- -----..A.. CA.T.T
	Z82r82	....TGGGATTT AACG...TC. C.G.A.---- ---------- ---------- ---------- -----CAA.. .....
	Z82r83	....TGGGATTT AACA..... ....A..A.A CTC.G----- ---------- ---------- -----.A.AG ....G.
	Z82r84	....TGGGATTT AACAA..... .....C A.G..----- ---------- ---------- --------.. .A..G.
	Z82r85	....TGGGGTTT AACA..... .....C A....----- ---------- ---------- --------.. .A....
	Z82r86	....G..ATC .T..... ....------ ...G..... ....A.... ---------- ----------.G .....
7	Z82r91	....TGGGAGTT AACA..... .....A CTC.G----- ---------- ---------- -----GA.A. ....G.
	Z82r92	....TGGGATTC AACA..... C.G...---- ---------- ---------- ---------- ------..A.G .....
	Z82r94	....C A.-------- ---------- ---------- -----.AA.. .....
	Z82r95	....G..ATT .T..... ....------ ...G...T.. ....A.... ---------- ----------..... .....
	Z82r96	....C A.-------- ---------- ---------- -----.AA.. .....
	Z82r97	....G..ATT .T..... ....------ ...G...T.. ....A.... ---------- ----------.G .....
	Z82r98	....C A...A----- ---------- ---------- -----GAA.. .....
	Z82r100	....TGGGATTT AACA....C. .....C A.G..----- ---------- ---------- -----.AA.. .....
	Z82r101	....G..ATT .T..... ....------ ...G...T.. ....A.... ---------- ----------.G .....
	Z82r102	....TGGGATTT AACA....C. ....C....C A.G..----- ---------- ---------- -----.AA.. .....
	Z82r103	....TGGGATTT AACA....C. ....C....C A.G..----- ---------- ---------- -----.AA.. .....

large number of gaps needed to be introduced to preserve the alignment. For example, 12 gaps were needed for sequence 82r986 to preserve alignment with sequences like 82r9814 (see Figure 4.1 [b]). However, despite the frequent sequence changes, not one inactivation mutation was observed in a total of approximately 35,000 bases obtained directly from plasma viral RNAs.

4.3.4 Nucleotide Distances of the V3 and V4 Sequences Within and Between Samples Pairwise nucleotide sequence distances were estimated among the sequences presented in Figure 4.1, using program DNADIST implemented in the PHYLIP package. Average sequence diversities both within and between samples were estimated on the basis of the pairwise distances and the results are presented in Tables 4.2 and 4.3.

Examination of Table 4.2 reveals that the distances between the seroconversion V3 sequence and that of the subsequent samples increased from 4.55% at year 3 (1987) to 8.02% at year 6 (1990), suggesting that later sequences diverge continually from the earliest sequence (see Table 4.2). However, comparison of nucleotide distances between years shows that they were not additive. Thus the mean nucleotide distance between samples from years 4 (1988) and 5 (1989) is 4.06% and that between years 5 (1989) and 6 (1990) was 5.21%. The mean distance between samples from years 4 (1988) and 6 (1990), however, is not (4.06 + 5.21 = 9.27%) but 5.14% (see Table 4.2). This non-additivity can be explained by the maximum likelihood phylogeny where two lineages are shown to split off from each other at year 3 (1987) and each fluctuates in frequency while continuing to diverge after that period (see

Table 4.2 Mean Pairwise Nucleotide Distances in the V3 Region of Patient 82

Sample[a]	Year 0	Year 3	Year 4	Year 5	Year 6	Year 7
Year 0	*0.0000*[b]					
Year 3	0.0455	*0.0326*				
Year 4	0.0648	0.0460	*0.0360*			
Year 5	0.0683	0.0510	0.0406	*0.0250*		
Year 6	0.0802	0.0633	0.0514	0.0521	*0.0575*	
Year 7	0.0796	0.0616	0.0548	0.0466	0.0577	*0.0467*

[a] Year 0 = 1984; year 3 = 1987; year 4 = 1988; year 5 = 1989; year 6 = 1990 and year 7 = 1991.

[b] Distances were estimated using program DNADIST implemented in PHYLIP package. Intra-sample distances are presented in *italics* on the diagonal. Mean distances from all pairwise inter-sample comparisons are presented below the diagonal.

Table 4.3 Mean Pairwise Nucleotide Distances in the V4 Region of Patient 82

Sample[a]	Year 0	Year 3	Year 4	Year 5	Year 6	Year 7
Year 0	*0.0726*[b]					
Year 3	0.2230	*0.0506*				
Year 4	0.1593	0.2984	*0.1974*			
Year 5	0.1355	0.4209	0.1471	*0.0172*		
Year 6	0.2243	0.5107	0.2508	0.1417	*0.2393*	
Year 7	0.2038	0.3775	0.2425	0.1908	0.2853	*0.3151*

[a] Year 0 = 1984; year 3 = 1987; year 4 = 1988; year 5 = 1989; year 6 = 1990 and year 7 = 1991.

[b] Distances were estimated using program DNADIST implemented in PHYLIP package. Intra-sample distances are presented in *italics* on the diagonal. Mean distances from all pairwise inter-sample comparisons are presented below diagonal.

section 4.3.5).

One of the most striking findings in this study is the homogenous nature of the viral population of V3 sequences at the time of seroconversion. Twelve sequences obtained are identical in both the nucleotide and amino acid levels. This contrasts strongly to what was observed in the later stage of infection where a substantial amount of sequence variation has been observed. A similar phenomenon has also been observed for another 4 samples collected prior to seroconversion (see Chapter 5).

In contrast to the between samples distances, the intra-sample distances do not show an persistent increase with time. Apart from that at seroconversion, the average intra-sample distance is approximately 3.9%. However, it is interesting to note that in year 5 (1989) the intra-sample distance was 2.05% and thus much lower than those of previous and subsequent years, suggesting the presence of a homogenous viral population with one type of virus variant dominating at that sampling time point.

The mean pairwise nucleotide distances in the V4 region, both between and within samples, were considerably higher than those of the V3 sequences in all but one case, suggesting the V4 region is under less constraint compared to the V3 region (see Table 4.3). The exceptionally low intra-sample distance in the V4 region observed at year 5 (1989) exactly corresponded to the lowest intra-sample distance of the V3 region illustrated above (Figure 4.1 and Tables 4.2 and 4.3), indicating that a homogenous viral population bearing very similar V3 and V4 sequences dominated the plasma population at the sampling period. Different from what was observed in the V3 region, the mean pairwise inter-sample nucleotide distances of the V4 region

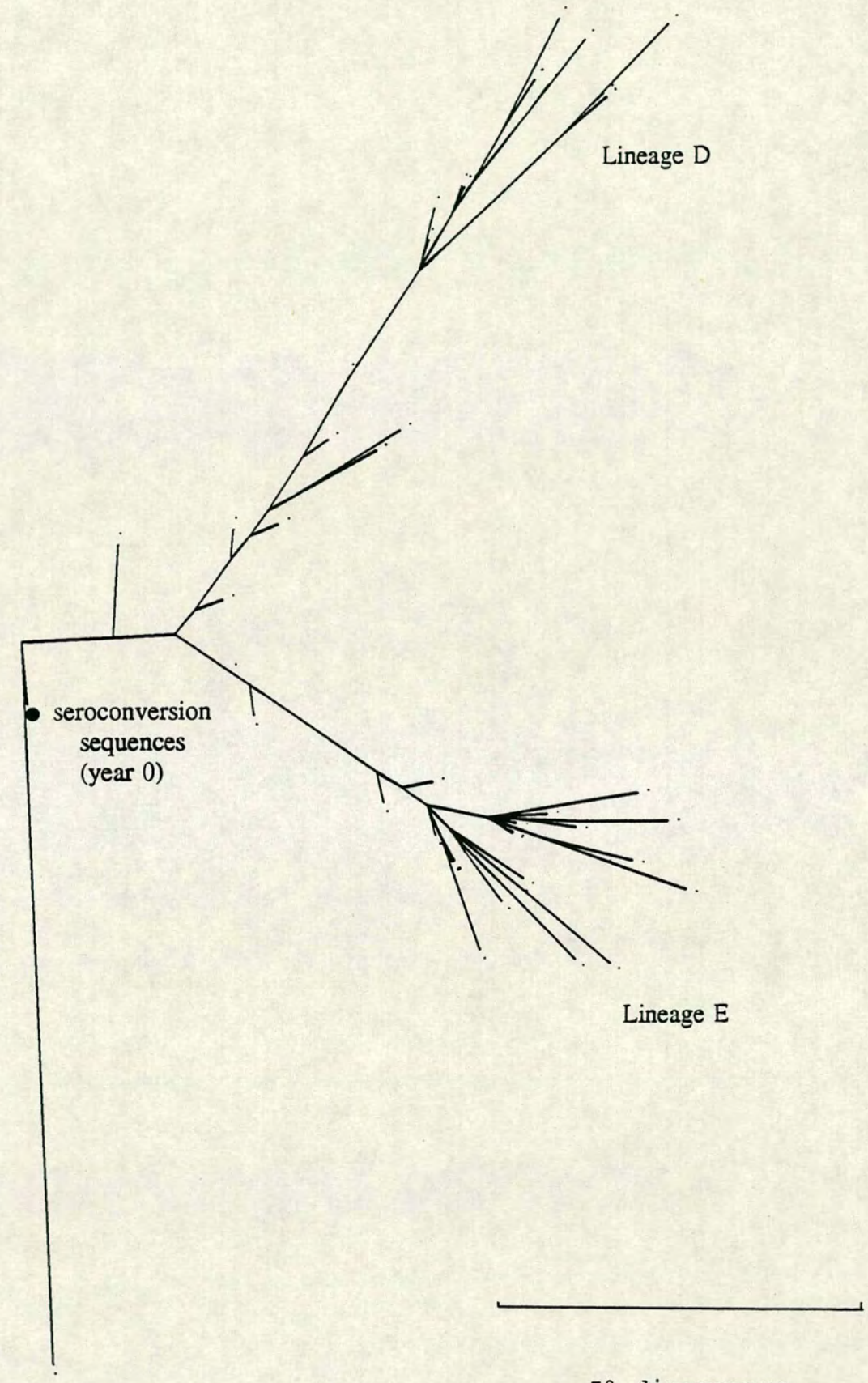
do not show a persistent increase over time. Between year 0 and year 3, the mean inter-sample distance is 22.3% whereas between year 0 and year 5, it decreases to 13.6%. However, it increases again between the year 0 and 6 and is comparable to that between year 0 and year 3 (see Table 4.3).

4.3.5 Phylogenetic Relationships among the V3 and V4 Nucleotide Sequences

Phylogenetic relationships among HIV sequences were estimated by the programs DNAML (DNA Maximum Likelihood) and NEIGHBOR (neighbor-joining) implemented in the PHYLIP package (version 3.4). Phylogenetic trees were constructed by Dr. Edward Holmes (Division of Biological Sciences, I.C.A.P.B., University of Edinburgh).

Phylogenetic analysis of the sequences obtained the from envelope genes of samples of patient 82 revealed that there are complex evolutionary patterns in the V3 and V4 hypervariable regions. For the V3 region, all the later sequences can be explained as progeny of the sequences found at seroconversion. At year 3 (1987), several equally distinct evolutionary lineages could be identified, but only two of them subsequently reached high frequency in the plasma population. This striking result was repeatedly obtained using different phylogenetic approaches such as Neighbor-Joining and Maximum Parsimony. Figure 4.2 depicts the global maximum likelihood tree for 89 V3 nucleotide sequences of plasma virions from patient 82 (courtesy of Dr. Edward Holmes, Division of Biological Sciences, I.C.A.P.B., University of Edinburgh). Individual sequences are presented by a dot, and a single sequence from $HIV_{HTLVIIIB}$ clone HXB2 was used as outgroup sequence (root sequence). The two

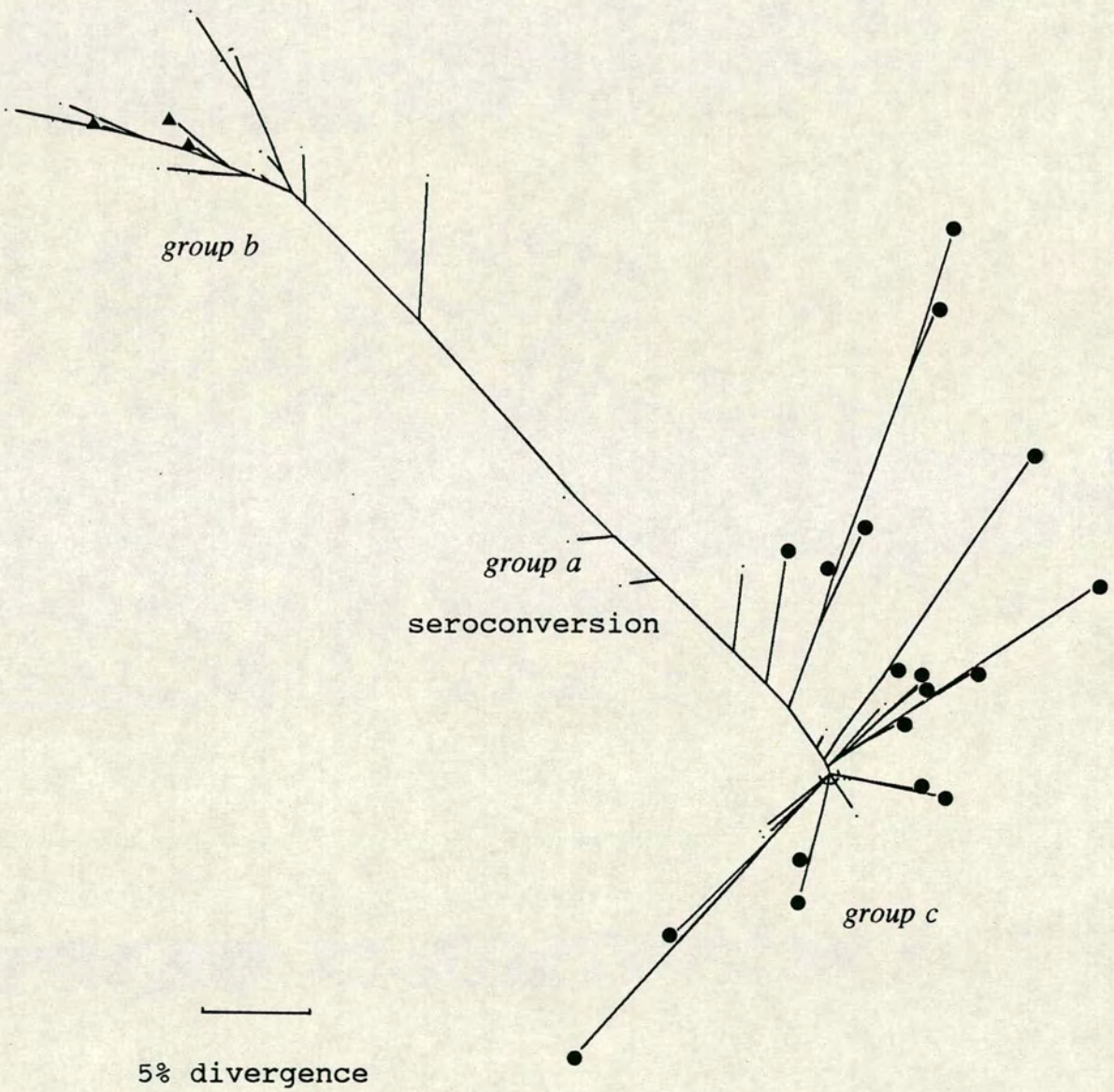
Figure 4.2. Maximum likelihood tree depicting evolutionary relationships among V3 sequences presented in Figure 4.1 (a). The tree is rooted by the HXB2 outgroup sequence. The seroconversion sequences (labelled as solid circle) are closest to the root sequence and therefore are suggested to be the ancestors of subsequent sequences. The major division into two main lineages is clearly seen and they are indicated as lineage D or E respectively. Each dot at the tips of the lineages represent a single individual sequence and all the branches are drawn to scale.


Lineage D
seroconversion
sequences
(year 0)
Lineage E
5% divergence
HXB2 (USA outgroup)

major lineages which were identified at year 3 and subsequently reached high frequency in the plasma population are indicated by Lineage D and E. Twelve identical V3 sequences found at seroconversion are closest to the root sequence and have been shown by different phylogenetic trees (constructed by Maximum Likelihood and Neighbor-Joining Methods) to be the ancestors of all subsequent sequences. Finally, all branch lengths of the tree are drawn to scale, which permits an assessment of the relative amounts of evolutionary change along different lineages.

The phylogenetic relationships among the 114 V4 viral RNA sequences of patient 82 are depicted in Figure 4.3. As for V3 sequences, individual V4 sequences are also represented by a dot at the tip of the branches (see Figure 4.3). All the branches are drawn to scale which enable the assessment of relative amount of evolutionary relatedness among these sequences. There is no outgroup sequence for the V4 tree because of the impossibility of sequence alignment with other published sequences. Examination of Figure 4.3 reveals that all the V4 sequences can be clustered into three main groups: *group a* includes the seroconversion sequences only (1984); *group b* sequences are those found frequently in year 4 (1988); and *group c* comprised the majority of sequences obtained in year 5 (1989) and onwards. The sequences from *group a* are likely to be the ancestors of those in *group b* and *group c* because of the availability of information on the order in which these sequences were found. As it is unlikely that sequences found at later stages of the infection could be the ancestors of sequences found at the first stage of HIV infection. At the same time, the branching pattern of the tree suggests that the sequences of *group b* and *c* may diverge independently from *group a.* Although the majority of sequences

Figure 4.3. Phylogenetic relationships among 114 V4 viral RNA sequences presented in Figure 4.1(b). As for the V3 sequences, individual V4 sequences are also represented by a dot at the tip of all the branches. All the branches are drawn to scale. Three sequence groups identified are labelled as *group a, group b* and *group c*, respectively. Sequences in subgroup *group c2* identified within *group c* sequences are indicated as solid circles, and they are frequently found in years 6 and 7. Sequences labelled as solid triangle in *group b* are those detected in years 6 and 7 after their disappearance in plasma population for more than two years (see section 4.3.11).


group b
group a
seroconversion
group c
5% divergence

in *group c* are found at year 5 (1989) and onwards, the mean inter-group nucleotide sequence distance between *group a* and *c* is (11.0%) no bigger than that between *group a* and *b* (18.6%). This finding suggests that succeeding sequences in *group c* may have evolved independently from those in *group a* and are not necessarily derived directly from those in *group b*. Interestingly, when sequences in *group c* were examined in more detail, two subgroups could be identified. One, designated as *group c1*, includes approximately 57 sequences mainly found at year 5 (1989) locating at the core of the branching pattern of *group c* sequences. The other, named as *group c2*, comprises the sequences stretching out from the *group c1* sequences and the majority of them are found in years 6 (1990) and 7 (1991) (labelled as solid circle in Figure 4.3). The close phylogenetic relationships between the sequences of *group c1* and *c2* indicates that the majority of sequences found in years 6 and 7 are directly derived from those found in year 5, although a couple of sequences found in year 6 and 7 also group with *group b* sequences (labelled as solid triangle in Figure 4.3).

4.3.6 Amino Acid Sequence Variation in the V3 Region of Plasma Viral RNAs

The translated amino acid sequences of the 89 V3 nucleotide sequences are presented in Figures 4.4. As can be seen from the Figure 4.4, the deduced amino acid sequences obtained over this study period showed a high degree of sequence variation. Many of the variants found in the V3 loop region of 245 North American isolates (LaRosa *et al.*, 1990) can also be found in this single patient. However, the distribution of amino acid replacements is not constant across the V3 region. If the deduced V3 region is further subdivided into three regions; the 35 amino acid V3 loop itself

Figure 4.4. Deduced amino acid sequences of the V3 region in patient 82. These amino acid sequences are deduced from nucleotide sequences presented in Figure 4.1 (a). Dots denote to the identical residues shown above. Hypervariable sites within the V3 loop are indicated as asterisk. Amino acid positions are numbered according to HIV_{HXB2}. Two cystein residues for the V3 loop disulphide bridge bond are at positions 296 and 330. Potential N-linked glycosylation sites are indicated as '#'. As before, year 0 = 1984; year 3 = 1987; year 4 = 1988; year 5 = 1989; year 6 = 1990 and year 7 = 1991.

Year	Sequences	277	296	* *	* *	* **	*	330	336
	Z82r001	FTDNAKTIIV	QLKESVEI#C	TRPN#NTRKS	IHIGPGRAFY	TTGEIIGDIR	QAHC#LSRAK		
	Z82r00a	....	....	....	....	....	....		
	Z82r00b	....	....	....	....	....	....		
	Z82r00c	....	....	....	....	....	....		
	Z82r00d	....	....	....	....	....	....		
0	Z82r00e	....	....	....	....	....	....		
	Z82r00f	....	....	....	....	....	....		
	Z82r00g	....	....	....	....	....	....		
	Z82r00h	....	....	....	....	....	....		
	Z82r001	....	....	....	....	....	....		
	Z82r00j	....	....	....	....	....	....		
	Z82r00k	....	....	....	....	....	....		
	Z82r003	..N.....	..#.....	....	.P......	...Q....	....		
	Z82r004	..N.....	..#.....	....	.....G...	...Q....	....		
	Z82r005	.A......	..#.....	....	.....G...	...Q....	.....I....		
	Z82r007	.A......	..#G....	....	.....G...	...Q....	....		
	Z82r008	....	....	....	....	...D.T..	....		
	Z82r009	....	.....V...	....	.P......	...D.T..	....		
	Z82r01	....	....	....	.P......	...D....	....		
3	Z82r02	....	....	....	.P......	...Q....	....		
	Z82r03	....	.....V...	....	.....G...	...Q....	....		
	Z82r04	....	....	....	....	...D....	.....I....		
	Z82r05	....	..#...V...	.........G	.....S...	A..D....	....		
	Z82r06	....	.....V...	....	.P......	...Q....	....		
	Z82r07	....	..#...V...	....	....	...G....	.....I....		
	Z82r08	....	..#...V...	....	....	...G....	.....I....		
	Z82r09	....	.....V...	....	.P......	...Q....	....		
	Z82r10	....	..#...V...	....	.....V.	...Q....	....		
	Z82r11	....	..#...V...	....	....	...G....	.....I....		
	Z82r12	....	..#...V...	.........G	.....S...	A..G....	.....I....		
	Z82r13	....	..#...V...	.........G	.....V.	..EQ...N..	.....I....		
	Z82r15	....	..#...V...	.........G	.....S...	A..G....	.....I....		
4	Z82r16	....	..#...V...	.........R	.....V.	..EQ...N..	.....I....		
	Z82r17	....	..#...V.T.	.........R	.....V.	..EQ...N..	.....I....		
	Z82r18	....	..#...V...	.........G	.....S...	A..R....	.....I....		
	Z82r19	....	..#G..V...	.........R	.....V.	..EQ...N..	.....I....		
	Z82r20	....	..#...V...	.........R	.....V.	..EQ...N..	....		
	Z82r21	....	..#...V...	.........R	.....V.	..EQ...N..	....		
	Z82r22	....	..#...V...	.........G	.....S...	A..G....	.....I....		
	Z82r24	....	..#...V.T.	.........R	.Y.....V.	..EQ...N..	.....I....		
	Z82r25	....	..#...V...	.........G	.....S.V.	A..G....	.....I....		
	Z82r26	....	..#...V.T.	.........R	.Y....S.V.	..EQ...N..	.....I....		
	Z82r27	....	..#...V...	.........G	.....S...	A..G...N..	.....I....		
	Z82r28	....	..#...V...	.........G	.....S...	A..D....	.....I....		
	Z82r29	.A......	..#.P.V...	.........G	.....S...	A..G....	.....I....		
	Z82r30	.A......	..#..TV...	.........G	.....S...	A..G....	.....I....		
	Z82r31	.A......	..#..TV...	.........G	.....S...	A..G....	.....I....		
	Z82r32	....	..#...V...	.........G	.....S...	A..G....	.....I....		
5	Z82r33	L.........	..#...V...	.........G	.....S...	A..G....	.....I....		
	Z82r34	L.........	..#...V...	.........G	.....S...	A..G....	.....I....		
	Z82r35	....	..#..TV...	.........G	.....S...	A..G....	.....I....		
	Z82r37	L.........	..#...V...	.........G	.....S...	A..G....	.....I....		
	Z82r38	....	..#...V...	.........G	.....S...	A..G....	.....I....		
	Z82r39	....	..#...V...	.........G	.....S...	A..G....	.....I....		
	Z82r40	....	..#...V...	.........G	.....S...	A..G....	.....I....		
	Z82r41	....	..#...V...	.........G	.....S...	A..G....	.....I....		
	Z82r42	....	..#...V.T.	.........R	.Y.....V.	..EQ...N..	.....I....		
	Z82r43	....	..#...V...	.........G	.....S...	A..G....	.....I....		
	Z82r44	....	..#...V...	.........G	.....S...	A..G...N..	.....I....		
	Z82r45	....	..#...V...	.........G	.....S...	A..G....	.....I....		
	Z82r46	....	..#...V...	.........G	.....S...	A..G....	.....I....		
	Z82r70	....	..#...V...	.....Y...R	.G.....SV.	.AEQ...N..	.....I....		
	Z82r71	....	..#...V...	.........G	.....S...	A..D....	....		
	Z82r72	....	..#...V...	.........G	.....S.V.	A..D....	.....I....		
	Z82r75	....	..#...V...	.....Y...R	.G.....SV.	.AEQ...N..	.....I....		
	Z82r76	....	..#...V...	.........G	....	A..D....	.....I....		
	Z82r77	....T.....	..#...V...	.....Y...R	.G.....SV.	.AEQ...N..	.....I....		
	Z82r78	....T.....	..#...V...	.........G	.....V.	...D....	.....I....		
6	Z82r79	....T.....	..#...V.T.	.........R	.Y.....V.	..DQ....	.....I....		
	Z82r80	....	..#...V...	.........G	.....S...	A..D....	.....I....		
	Z82r81	....	..#A..V.T.	.........R	.Y.....V.	..DQ...N..	.....I....		
	Z82r82	....	..#...V...	.........G	.....S...	A..D....	.....I....		
	Z82r83	....T.....	..#...V.T.	.........R	.Y.....V.	..DQ...N..	.....I...T		
	Z82r84	....	..#A..V.T.	.........R	.Y.....V.	..DQ...N..	.....I...T		
	Z82r85	....	..#...V...	.........G	.....S...	A..D....	.....I....		
	Z82r86	....	..#...V...	.........G	.....S...	A..D....	.....I....		
	Z82r91	....T.....	..#K..V.T.	.........R	.Y.....V.	..DQ...N..	.....I....		
	Z82r92	....	..#...V...	.........G	.....S...	A..D....	.....I....		
	Z82r93	....	..#...V...	.........G	.....S...	A..D....	.....I....		
	Z82r94	....	..#...V...	.........G	.....S...	A..D....	.....I....		
	Z82r95	....	..#...V...	.........G	.....S...	A..D....	.....I....		
	Z82r96	....	..#...V...	.........G	.....S...	A..D....	.....I....		
7	Z82r97	....	..#...V...	.........G	.....S...	A..D....	.....I....		
	Z82r98	....	L.#K..V...	.........G	.....S...	A..D....	.....I....		
	Z82r99	....	..#...V...	.........G	.....S...	A..D....	.....I....		
	Z82r100	....T.....	..#...V...	.....Y...R	.G.....SV.	.AEQ...N..	.....I....		
	Z82r101	....	..#...V...	.........G	.....S...	A..D....	.....I....		
	Z82r102	....	..#...V...	....	....	...G....	.....I....		
	Z82r103	....	..#...V...	.........G	.....S...	A..D....	.....I....		

Year	Sequences	337 353
0	Z82r001	WTDTLRQIVM KLREQFG
	Z82r00a	....
	Z82r00b	....
	Z82r00c	....
	Z82r00d	....
	Z82r00e	....
	Z82r00f	....
	Z82r00g	....
	Z82r00h	....
	Z82r00i	....
	Z82r00j	....
	Z82r00k	....
3	Z82r003	.#E......I ..G....
	Z82r004	.#E......I ......E
	Z82r005	.#E..K...I ......E
	Z82r007	.#E..K...I ..GK...
	Z82r008	.#E..K...I ..G....
	Z82r009	.#E..K...I ..G....
	Z82r01	.#E......I ..GK...
	Z82r02	.#E......I ..G....
	Z82r03	.#E......I ..G....
	Z82r04	.#...GK... ..G....
	Z82r05	.#E..K...I ..G...R
	Z82r06	.#E......I ..G....
	Z82r07	.#E..K...I ......E
	Z82r08	.#E..K...I ......E
	Z82r09	.#E......I ..G....
4	Z82r10	.#E......I ..G....
	Z82r11	.#E..K...I ....
	Z82r12	.#E..K...I ......E
	Z82r13	.#E..K...I ......E
	Z82r15	.#E..K...I ......K
	Z82r16	.#E..K...I ......E
	Z82r17	.#E..K...I ......E
	Z82r18	.#E..K...I ......E
	Z82r19	.#E..K...I ......E
	Z82r20	.#E......I ..G....
	Z82r21	.#E......I ..G....
5	Z82r22	.#E..K...T ......K
	Z82r24	.#E..K...I ......E
	Z82r25	.#E..K...I ....
	Z82r26	.#E..K...I ......E
	Z82r27	.#E..K...I ....
	Z82r28	.#E..K...I ......E
	Z82r29	.#E..K...I ....
	Z82r30	.#E..K...I ....
	Z82r31	.#E..K...I ....
	Z82r32	.#E..K...I ......K
	Z82r33	.#E..K...I ....
	Z82r34	.#E..K...I ....
	Z82r35	.#E..K...I ......K
	Z82r37	.#E..K...T ......K
	Z82r38	.#E..K...I ....
	Z82r39	.#E..K...I ....
	Z82r40	.#E..K...I ....
	Z82r41	.#E..K...T ....
	Z82r42	.#E..K...I ......E
	Z82r43	.#E..K...I ....
	Z82r44	.#E..K...I ....
	Z82r45	.#E..K...I ....
	Z82r46	.#E..K...I ....
6	Z82r70	.#E..K...I ....
	Z82r71	.#E..K...I ....
	Z82r72	.#E..K...I ....
	Z82r75	.#E..K...L ......E
	Z82r76	.#E..K...I ....
	Z82r77	.#E..K...I ......D
	Z82r78	.#E..K...T ......D
	Z82r79	.#E..K...T ......K
	Z82r80	.#E..E...I ....
	Z82r81	.#E..K...T ....
	Z82r82	.#E..K...T ....
	Z82r83	.#...K...T ....
	Z82r84	.#...K...I ....
	Z82r85	.#K..E...I ....
	Z82r86	.#K..E...T ....
7	Z82r91	.#E..K...I ......K
	Z82r92	.#E..EK..I ....
	Z82r93	.#K..E...I ......E
	Z82r94	.#K..E...T ......E
	Z82r95	.#K..E...T ......E
	Z82r96	.#K..E...I ....
	Z82r97	.#K..E...T ......E
	Z82r98	.#E..K...I ....
	Z82r99	.#K..E...T ......E
	Z82r100	.#E......L ......K
	Z82r101	.#K..E...T ......E
	Z82r102	.#E..K...I ....
	Z82r103	.#K..E...T ....

(bounded by 2 cysteine residues at positions 296 and 330) and two flanking regions (a 20-residue 5' region and a 24-residue 3' region), the rate of amino acid replacement is much higher in the V3 loop than in the 5' and 3' regions flanking (Figure 4.4). In the entire sequence data, there were a total of 24 different V3 loop amino acid sequences. Based on their amino acid sequences on all the variable sites with the loop, these sequences can be further categorised into different sequence types, designated from A to F, which are presented in Table 4.4, together with the frequencies at which they are found in each year (courtesy of Dr. E. Holmes, Division of Biological Sciences, I.C.A.P.B., University of Edinburgh University). Each of these sequences can be seen as having one or more amino acid differences from that found at seroconversion (designated as sequence A). As shown in the Table 4.4, the amino acid replacements within the V3 loop are largely concentrated at positions 306, 308, 313, 315, 317, 319, 320, 324, some of which have been shown to be within the target sequence recognized by neutralizing antibodies (306, 308) (Looney *et al.*, 1988; Mckeating *et al.*, 1989) and cytotoxic T cell (Takahashi *et al.*, 1988), and responsible for conversion for one phenotype to another (De Jong *et al.*, 1992*a, b*; Fouchier *et al.*, 1992). At amino acid residue position 306, the most commonly found residue at year 3 was serine (S), subsequently it changed to arginine (R) at year 4 and in year 7 was glycine (G). Similarly at positions 319 and 320, at the time of seroconversion (at year 0), the respective amino acid sequences were glycine (G)-Glutamate (E), however, they changed to glycine (G)-glutamine(Q) at year 3; glutamate(E)-glutamine(Q) at year 4; glycine(G)-glycine(G) at year 5 and glycine(G)-aspartate(D) at year 6. The coincidence of the hypervariable nature and the high degree of antigenicity in the V3

Table 4.4. The 24 V3 loop amino acid sequences and their samples frequencies in the plasma. Only residues that differ from the seroconversion sequences (lineage A) are shown. As before, amino acid positions are numbered according to HIV_{HXB2}. Hypervariable sites within the V3 loop are indicated as asterisk. Potential N-linked glycosylation sites are indicated as '#'. The frequency of each sequence in each year are presented, with total number of nucleotide sequences obtained given at bottom. As before, year 0 = 1984; year 3 = 1987; year 4 = 1988; year 5 = 1989; year 6 = 1990 and year 7 = 1991.

Sequence genotype		Frequency in Year					
		0	3	4	5	6	7
	296 * * * * * ** * 330						
A	CTRPN#NTRKSIHIGPGRAFYTTGEIIGDIRQAHC	1.000					
B	.....#.....D.....		0.067				
C1	.....#.....P.....D.....		0.067				
C2	.....#.....P.....Q.....		0.267				
C3	.....#.....G.....Q.....		0.267				
C4	.....#.....P.....D.T.....		0.067				
C5	.....#.....D.T.....		0.067				
D1	.....#.....V....Q.....			0.091			
D2	.....#....R.....V...EQ...N.....			0.455			
D3	.....#....G.....V...EQ...N.....			0.091			
D4	.....#....R.Y.....V...EQ...N.....				0.087		
D5	.....#....R.Y....S.V...EQ...N.....				0.043		
D6	.....#....R.Y.....V...DQ.....					0.067	
D7	.....#....R.Y.....V...DQ...N.....					0.200	0.077
D8	.....#Y...R.G.....SV..AEQ...N.....					0.200	0.077
E1	.....#....G.....S...A..D.....		0.067		0.043	0.333	0.769
E2	.....#....G.....S...A..G.....			0.182	0.696		
E3	.....#....G.....S...A..R.....			0.091			
E4	.....#....G.....S.V.A..G.....				0.043		
E5	.....#....G.....S...A..G...N.....				0.087		
E6	.....#....G.....S.V.A..D.....					0.067	
E7	.....#....G.....A..D.....					0.067	
E8	.....#....G.....V....D.....					0.067	
F	.....#.....G.....		0.133	0.091			0.077
Total		12	15	11	23	15	11

region indeed support the notion that high rates of amino acid changes in this region may facilitate virus variants to escape from immune clearance (Albert *et al.*, 1990; Montefiori *et al.*, 1991).

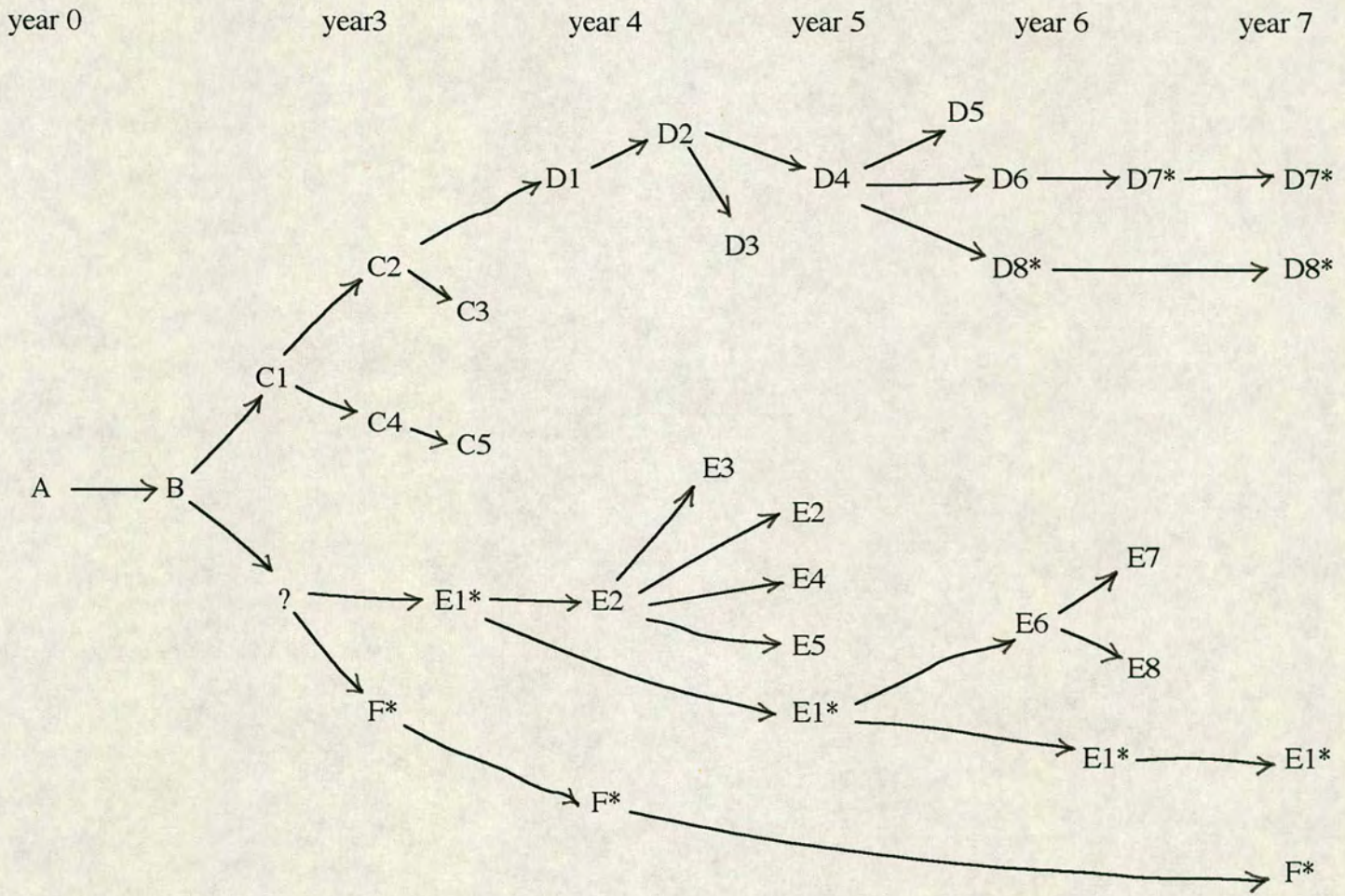
In the flanking regions of the V3 loop, several amino acid replacements (at position 289, 295 and 338) alter the potential sites for N-linked glycosylation (indicates as '#' in Figure 4.4). Take amino acid residue at the position 289 for example; at year 3, over 50% (8 out of 15) of viral RNA sequences were Lysine (K) and they all became Asparagine (N) in the subsequent years (Figure 4.4). Similarly, at position 295, some of the virus variants found after 3 years of infection had lost their potential glycosylation sites (Molecules Z82r17, Z82r24, Z82r26, Z82r42, Z82r79, Z82r81, Z82r83, Z82r84 and Z82r91, see Figure 4.4). The regaining and loss of these sites may also have direct effect on virus antigenicity and infectivity (Montefiori *et al.*, 1988; Hansen *et al.*, 1990).

Despite the hypervariability at some sites within the V3 loop, there are several segments, namely CTRPNNNTRK, IGPG and IRQAHC which constitute the 'framework' of the V3 loop, and show much less variability. The IGPG tetrapeptide, for instance, was present in all sequences obtained and so were the two cysteines (at position 296 and 330) for the V3 loop disulphide bridge bond. Moreover, the twelve sequences obtained from the plasma sample from patient 82 at seroconversion are identical among themselves at both nucleotide and amino acid levels, and the amino acid sequence is the same as the 'global consensus V3 sequence' identified by LaRosa *et al.*, from 245 North American isolates (LaRosa *et al.*, 1990). This finding contrasts strongly to what has been observed at the later stage of infection where the plasma

viral population is extremely heterogenous (Figure 4.4 and Table 4.4). The seroconversion sequences differed at many sites from viral RNA sequences identified after 3-years of infection, but was identical to three of the five proviral DNAs obtained from the same year (DNA sequence data was kindly provided by Dr. Peter Simmonds and Mrs Fiona McOmish, Simmonds *et al.*, 1991). Therefore, it appears that seroconversion sequences can persist much longer in proviral DNA than in viral RNA form. Similarly, V4 sequence type (*group a*) found only in the plasma sample at seroconversion were still present in proviral DNA 3 years later (Simmonds *et al.*, 1991).

4.3.7 Evolution of the V3 Loop If the phylogenetic relationships of the nucleotide sequences from the V3 region can be established, evolution of the antigenic structure of the V3 loop can be studied. This has been done by superimposing the V3 loop amino acid sequences onto a phylogenetic tree inferred from the nucleotide sequences using the DNAML program. The resulting evolutionary pattern relating the 24 different amino acid sequences is depicted in Figure 4.5 (courtesy of Dr. Edward Holmes, Division of Biological Sciences, I.C.A.P.B., University of Edinburgh). As shown in Figure 4.5, many amino acid changes in the V3 loop correspond to the evolutionary pattern observed at nucleotide level (see Figure 4.2). It is quite clear that evolutionary lineages D and E are two distinct lineages derived from early sequences which subsequently reached high frequency (Table 4.4). These two lineages dominate the whole viral population and they seem to do so by alternating in abundance from year 4 and onwards (Table 4.4). This is reflected by changes in their frequency within

Figure 4.5. Evolutionary process relating the defined 24 amino acid sequences found in the V3 loop. Lineages are designated by letters A to F. Proposed evolutionary relationships are indicated by arrows. Lineages that persist through the study are labelled by asterisk. Time scale is shown along the top (courtesy of Dr. Edward Holmes).


year 0
year3
year 4
year 5
year 6
year 7
A
B
C1
C2
C3
C4
C5
D1
D2
D3
D4
D5
D6
D7*
D7*
D8*
D8*
?
E1*
E2
E3
E2
E4
E5
E6
E7
E8
E1*
E1*
E1*
F*
F*
F*

the whole viral population. At year 4, for instance, lineage D composed nearly 50% of the whole viral population, whereas lineage E composed less than 20%. However, at year 5, more than 70% of virus variants were from lineage E while less than 15% were from lineage D. At year 6, lineages D and E were at comparable levels and each consists of roughly 50% of the viral population. At year 7, however, lineage E frequency rises once again and nearly 80% of virus variants present in the plasma were from this lineage (see Table 4.4).

4.3.8 Amino Acid Sequence Variation in the V4 Region of Plasma Viral RNAs

The deduced amino acid sequences of the 114 V4 nucleotide sequences are presented in Figure 4.6. Of these sequences, three groups or lineages, designated as A, B and C, have been identified based on their nucleotide sequence and sequence length. These amino acid sequences are presented in Table 4.5, together with their frequencies found each year. The length of the consensus sequence in Group A is 22 amino acids between the highly conserved flanking segments FNSTW and ITLPCR, whereas that of Group B and Group C are 17 and 18 amino acids long respectively. The consensus sequences of these three groups are clearly distinct from each other, whereas the individual sequences within each group are very similar, although some sequences are shorter or longer. Similarly, the potential sites for N-linked glycosylation (indicated as '#') in this region are also different between these identified sequence groups (see Figure 4.6). Interestingly, the sequences obtained at the years 0 (1984), 3 (1987), 4 (1988) and 5 (1989) fit these defined groups much better than those obtained in the later samples (see Figure 4.6). The burst of length and sequence variation in the years

Figure 4.6 Deduced amino acid sequences of the V4 region in patient 82. These amino acid sequences are deduced from nucleotide sequences presented in Figure 4.1 (b). Gaps were introduced to preserve the sequence alignment and indicated by dashes. Dots denote to the identical residues shown above. Potential N-linked glycosylation sites are indicated as '#'. Amino acid positions are numbered according to HIV_{HXB2}. As before, year 0 = 1984; year 3 = 1987; year 4 = 1988; year 5 = 1989; year 6 = 1990 and year 7 = 1991.

Year	Sequences	389 424
0	82r002	F#STW #STQL#STW#STQL#SAG##TEE# ITLPCR
	82r005	.#... #STQL#STW#---------#TEE# .....
	82r003	.#... #SIQF#STW#STQL#S-----EE# .....
	82r004	.#... #STQF#STW#STQL#SA--RTEE# .....
	82r00a	.#... #STQF#STW#STQL#SA--RTEE# .....
	82r00b	.#... #STQF#STW#STQL#SA--RTEE# .....
	82r00c	.#... #STQF#STW#STQL#SA--RTEE# .....
	82r00d	.#... #STQF#STW#STQL#SA--RTEE# .....
	82r00e	.#... #STQF#STW#STQL#SA--RTEE# .....
3	82r5571	.#... #--YS#GTW#-----STQH#TGE# .....
	82r5572	.#... #--YY#GTW#-----STQH#TEE# .....
	82r5575	.#... #--YS#GTWI-----STQH#TEE# .....
	82r5577	.#... #--YS#DTWS-----STQH#TEE# .....
	82r5579	.#... #--YS#G#WS-----STQH#TEE# .....
	82r5576	.#... #--YS#G#WS-----STQH#TEE# .....
	82r5573	.#... #--YS#G#WS-----STQH#TEE# .....
	82r5578	.#... #--YS#G#WS-----STQH#TEE# .....
	82r5580	.#... #--YS#G#WS-----STQH#TEE# .....
	Z82r01	.#... #--YF#GTW#-----STQH#TGE# .....
	Z82r02	.#... #--YS#G#WS-----STQH#TEE# .....
	Z82r03	.#... #--YS#G#WT-----STQH#TGE# .....
	Z82r04	.#... #--YS#DTWS-----STQH#TEE# .....
	Z82r05	.#... D--YS#STW#-----SAQI#TEE# .....
	Z82r06	.#... #--YS#GTWS-----STQH#TEE# .....
	Z82r07	.#... #--YS#GTWS-----STQH#TEE# .....
	Z82r08	.#... #--YS#GTWI-----STQH#TEE# .....
	Z82r09	.#... #--YS#G#WS-----STQH#TEE# .....
4	82r122	.#... #W#YS#G#WS-----STQH#TGE# .....
	82r1210	.#... S--YS#G#WS-----STQH#TGE# .....
	82r125	.#... S--YS#GTW#-----STQH#TGE# .....
	82r1212	.#... -----#STWDLTQL#STQ#K-EE# .....
	82r1213	.#... -----#STWDLTQL#STQ#K-EE# .....
	82r1214	.#... -----#STWDLTQL#STQ#K-EE# .....
	82r1223	.#... -----#STWDLTQL#STQ#K-EE# .....
	82r126	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r10	.#... T--YS#DTW#-----STQH#TGE# .....
	Z82r11	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r12	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r13	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r15	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r16	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r17	.#... -----#STWDLTQL#STQ#K-GE# .....
	Z82r18	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r19	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r20	.#... S--YS#GTW#-----STQH#TGE# .....
	Z82r21	.#... S--YS#GTW#-----STQH#TGE# .....
	82r821	.#... -----#STWDLTQL#SIQ#R-EE# .....
	82r825	.#... -----#STWDLTQL#STQ#K-EE# .....
	82r8215	.#... -----#STWDLTQL#STQ#K-EE# .....
	82r1091	.#... -----#STWDLTQL#STQ#K-EE# .....
	82r1092	.#... -----#STWDLTQL#STQ#K-EE# .....
	82r1094	.#... -----#STWDLTQL#STQ#K-EE# .....
	82r1095	.#... -----#STWDLTQL#STQ#K-EE# .....
	82r1097	.#... -----#STWDLTQL#STQ#K-EE# .....
	82r1098	.#... -----#STWDLTQL#STQ#K-EE# .....

	82r1096	.#... -----#STWDLTQL#STQ#K-EE# .....
	82r1099	.#... -----#STWDLTQL#STQ#K-EE# .....
	82r1093	.#... -----#STWDLTQL#STQ#K-EE# .....
	82r1100	.#... -----#STWDLTQL#STQ#K-EE# .....
	82r984	.#... -----#STWDLTQL#GTQ#K-EE# .....
	82r9814	.#... -----#STWDLTQL#STQ#K-EE# .....
	82r986	.#... -----#STWDLTQ----Q#K-EE# .....
	82r9810	.#... -----#STWDLTQL#STQDK-KD# .....
	82r988	.#... -----#STWDLTQL#STQDK-KD# .....
	82r9817	.#... -----#STWDL#S---TQ#K-EE# .....
	82r9820	.#... -----#STWDLTQP#STQ#K-EE# .....
	Z82r22	.#... -----#STWDLTQL#STQ#K-EE# .....
5	Z82r24	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r25	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r26	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r27	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r28	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r29	.#... -----#STWDLT-L#STQ#K-EE# .....
	Z82r30	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r31	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r32	.#... -----#STWDLTQL#YSQ#K-EE# .....
	Z82r33	.#... -----#STWDLTQL#RAQ#K-EE# .....
	Z82r34	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r35	.#... -----#STWDLTQLGSTQ#K-EE# .....
	Z82r37	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r38	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r39	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r40	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r41	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r42	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r43	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r44	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r45	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r46	.#... -----#STWDLTQL#STQ#K-EE# .....
	Z82r70	.#... -----#STWDSTQLQST----EE# .....
	Z82r71	.#... -----#STQL#STQ-----K-EE# .....
	Z82r72	.#... -----#STWDLT#LDG#Q--EGE# .....
	Z82r76	.#... -----#STGDLTQL#STQ-K-EG# .....
	Z82r77	.#... -----#STWDLTQL#STQDK-G-# .....
	Z82r78	.#... -----#STWDLTQL#GTQDK-GE# .....
6	Z82r79	.#... -----#STWDLTQL###TQ#REE# .....
	Z82r80	.#... ---------#-----STQH#TGE# .....
	Z82r81	.#... -----#STWELTQL#STQDT-D## .....
	Z82r82	.#... -----#STWDLTQ---SQ#Q-EE# .....
	Z82r83	.#... -----#STWDLTQL###TQ#REE# .....
	Z82r84	.#... -----#STWDLTKL#STQD--EK# .....
	Z82r85	.#... -----#STWGLTQL#STQ#--EK# .....
	Z82r86	.#... S--YP#GTW#-----STQH#TGE# .....
	Z82r91	.#... -----#STWELTQL#STTQD-KE# .....
	Z82r92	.#... -----#STWDSTQLQST----GE# .....
	Z82r94	.#... -----#STQL#STQ-----K-EE# .....
	Z82r95	.#... S--YS#GIW#-----STQH#TEE# .....
	Z82r96	.#... -----#STQL#STQ-----K-EE# .....
7	Z82r97	.#... S--YS#GIW#-----STQH#TGE# .....
	Z82r98	.#... -----#STQL#STQ-----KEEE# .....
	Z82r100	.#... -----#STWDLTQP#STQDK-EE# .....
	Z82r101	.#... S--YS#GIW#-----STQH#TGE# .....
	Z82r102	.#... -----#STWDLTQP#TTQDK-EE# .....
	Z82r103	.#... -----#STWDLTQP#TTQDK-EE# .....

Table 4.5. The 16 sequence genotypes of the V4 region and their samples frequencies in the plasma over the study period. Only residues that differ from A1 sequences are shown. Gaps were introduced to preserve the alignment and indicated by dashes. Dots denote to the identical residues shown at the top (A1 sequences). Potential N-linked glycosylation sites are indicated as '#'. The frequency of each sequence in each year are presented, with total number of nucleotide sequences obtained given at bottom. As before, year 0 = 1984; year 3 = 1987; year 4 = 1988; year 5 = 1989; year 6 = 1990 and year 7 = 1991.

	Sequence genotype	Frequency in Year					
		0	3	4	5	6	7
A1	F#STW #STQF#STW#STQLNSA--RTEE# ITLPCR	0.667					
A2	.#... #STQL#STW#STQLNSAG#NTEE#	0.111					
A3	.#... #SIQF#STW#STQLNS-----EE#	0.111					
A4	.#... #STQL#STW#---------NTEE#	0.111					
B1	.#... N--YSNG#WSSTQH-----NTEE#		1.000	0.263		0.071	0.272
B2	.#... NWNYSNG#WSSTQH-----NTEE#			0.052			
B3	.#... ---------#STQH-----NTGE#					0.071	
C1	.#... #STWDLTQL#STQ------NKEE#			0.684	0.953	0.142	0.363
C2	.#... #S----TQL#STQ-------KEE#					0.071	0.272
C3	.#... #STWDL---#STQ------NKEE#				0.023		
C4	.#... #STWDLTQQ----------NKEE#				0.023		
C5	.#... #STWDLTQLN#NTQ-----NREE#					0.142	
C6	.#... #STWDSTQLQST--------GEE#						0.091
C7	.#... #STWDLTNLDGNQ-------EGE#					0.357	
C8	.#... #STWDKTQSQ---------NQEE#					0.071	
C9	.#... #STWDSTQLQST---------EE#					0.071	
Total		9	18	19	43	14	11

6 (1990) and 7 (1991) is obviously the reason that the sequences are distinct from the defined group consensus sequences. In consequence, shorter and longer forms were both frequently found. For example, the Group B sequences were exclusively 17 amino acids long between the relatively conserved flanking regions FNSTW and ITLPCR at years 3 and 4. However, at year 6, a Group B sequence (Z82r80) became 10 amino acids long (see Figure 4.6). Similarly, the length of the Group C type sequence was 18 amino acids in length between segments FNSTW and ITLPCR before the year 6 (1990). However at years 6 (1990) and 7 (1991), both longer and shorter forms of Group C sequences were frequently found (see Figure 4.6). Sequences Z82r71, Z82r94 and Z82r96 are 13 residues in length between segments FNSTW and ITLPCR whereas sequences Z82r79 and Z82r83 have 19 residues (see Figure 4.6). The length of Group A sequences are somewhat variable during the early stage of infection. Four length types in total were found shortly after seroconversion with one type domination (see Figure 4.6).

4.3.9 The Evolution of the V4 Region As for the V3 region, the 114 V4 amino acid sequences were superimposed on to the phylogenetic tree defined from 114 V4 nucleotide sequences (see section 4.3.7) to study the relationships between grouping pattern defined at both amino acid and nucleotide levels. Not unexpectedly, the sequence groups defined at the amino acid level correlated very well with the lineages identified by phylogenetic analysis; with amino acid sequences of Group A corresponding to nucleotide sequences of *group a*, Group B to *group b* and Group C to *group c*, indicating that many amino acid changes in the V4 region correspond to

the evolutionary pattern observed at nucleotide level. The consequent evolutionary pattern of the various genotypes between and within defined groups is depicted in Figure 4.7. In contrast to what was observed in the V3 region, the evolutionary pattern in the V4 region is more clear and can be assessed more easily. First of all, two lineages, derived from group a sequences, can be clearly seen and they are labelled as lineages B and C. Lineage B sequences are very abundant during the third year after seroconversion (1987) and were subsequently almost undetectable for more than two years. Lineage C sequences consist primarily of those found at years 5, 6 and 7. Although these sequences appeared a year after of those in lineage B, the evolutionary distance between lineage C and A sequences is no bigger than that between lineage B and A sequences (see Table 4.3). Therefore, lineage B and C are likely to derive independently from lineage A, which once again suggests a discontinuous evolutionary process. It is quite clear that C2-C8 genotype sequences are derived from C1 whereas B2 and B3 are from B1.

4.3.10 Distinct but Related Plasma Viral and Lymphocyte-associated Proviral Populations The rapid sequence turnover observed in the plasma viral population can also be identified coincidently in the lymphocyte-associated proviral population of the same patient, although this change turns out to lag behind that of the plasma viral RNA population. Figure 4.8, depicts the changes in the type and frequency of proviral DNA V4 sequences present in PBMCs over the 5 years post-infection (proviral DNA sequences are kindly provided by Dr. Peter Simmonds and Mrs. Fiona McOmish, Department of Medical Microbiology, University of Edinburgh) and compared with

Figure 4.7. Evolutionary process relating the defined 16 amino acid sequences found in the V4 region. Lineages are designated by letters A to C. Proposed evolutionary relationships are indicated by arrows. Lineages that persist through the study are labelled by asterisk. Time scale is shown along the top.

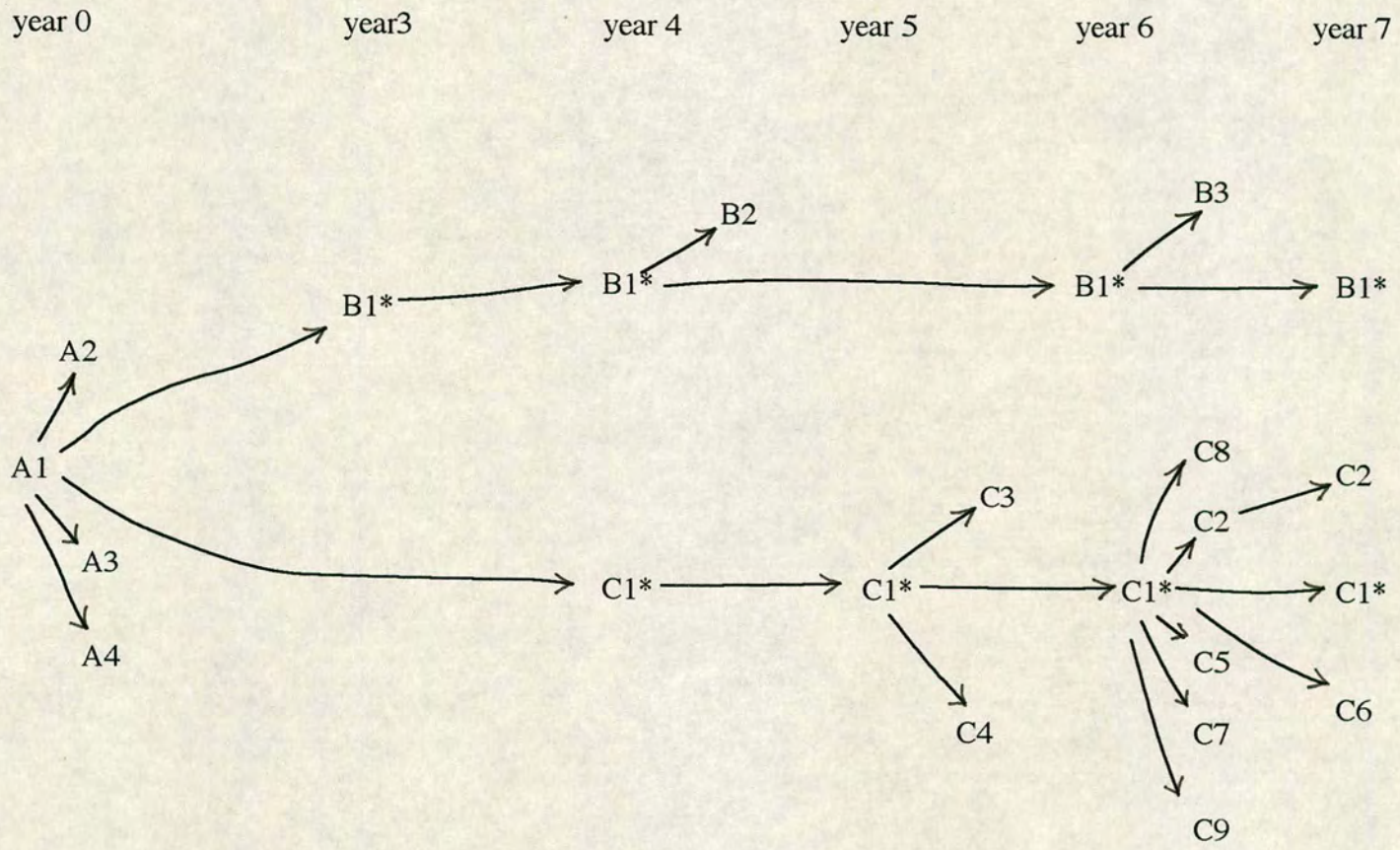
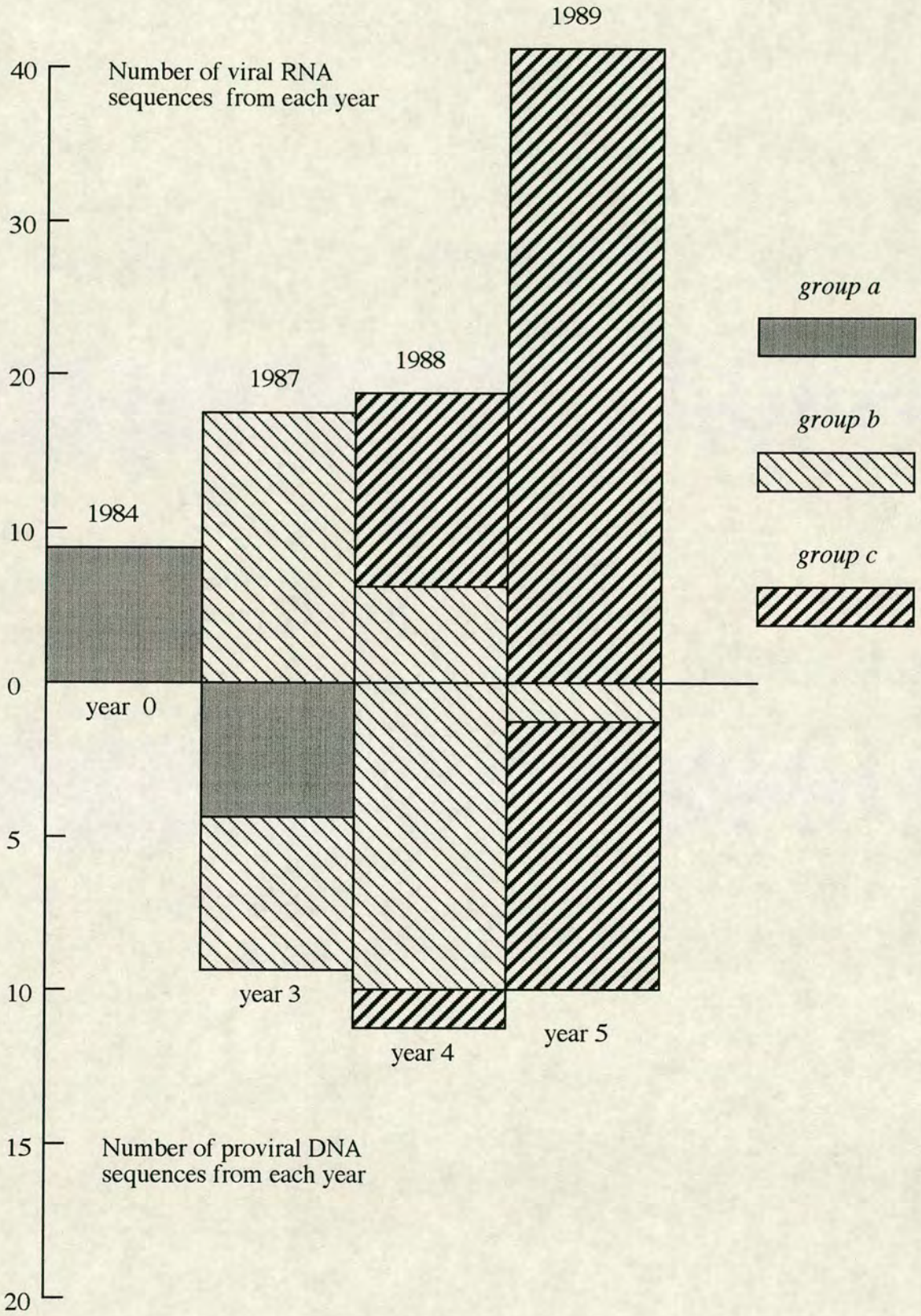

year 0
year3
year 4
year 5
year 6
year 7
A1
A2
A3
A4
B1*
B1*
B2
B1*
B3
B1*
C1*
C1*
C3
C4
C1*
C8
C2
C2
C1*
C5
C7
C9
C6

Figure 4.8. Frequency of detection of *group a, b* and *c* sequences in sequential PBMCs and plasma samples from patient 82. Viral RNA sequences from plasma are shown above the *x* axis, and proviral DNA sequences from PBMCs are shown below the *x* axis.


1989
40
Number of viral RNA
sequences from each year
30
group a
1988
20
1987
group b
1984
10
group c
0
year 0
5
10
year 3
year 5
year 4
15
Number of proviral DNA
sequences from each year
20

those found in the plasma viral population (Simmonds *et al.*, 1991). At year 3 (1987), for example, both *group a* and *group b* sequences were found in the proviral DNA population whereas in the viral population only *group b* were detected. However, in the following year (1988), virtually all proviral DNA sequences were in *group b* whereas the majority of the viral RNA sequences (72%) became *group c* sequences. At year 5 (1989) when the complete replacement by *group c* sequences had finished in the plasma viral population, the process of this replacement was still going on in the proviral DNA population.

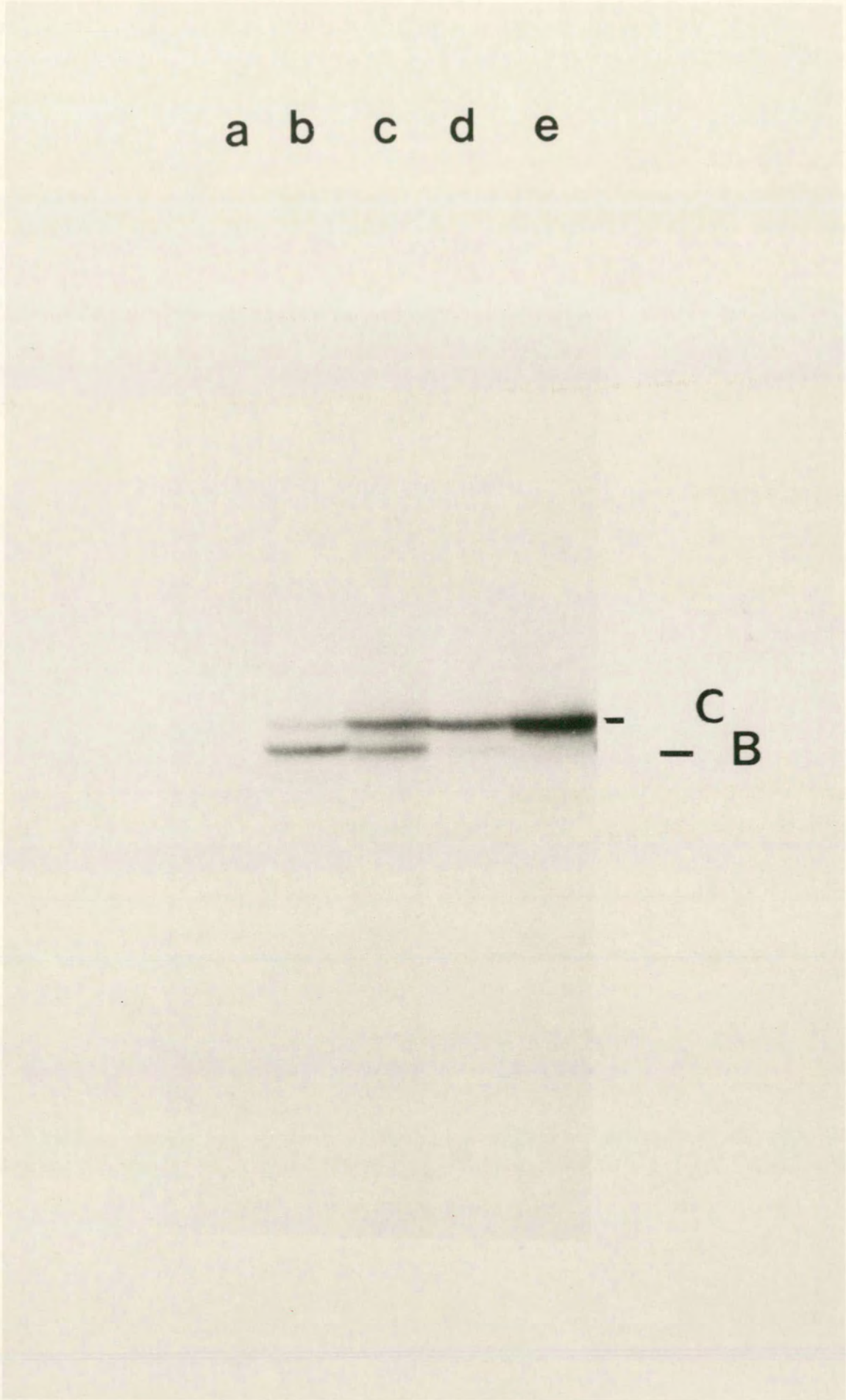
Having studied the similarities and the differences between the plasma viral and lymphocyte-associated proviral populations, a general trend was found that RNA sequences turned over more rapidly than the corresponding DNA sequences, hence the new virus variants appeared and became dominant firstly in the plasma rather than in the PBMCs populations. For example, the seroconversion *group a* sequence was completely replaced in plasma at year 3 (1987) yet formed a substantial proportion of the sequences in PBMCs at that time. Similarly, the difference in the relative numbers of *group b* and *group c* sequences in the sample collected at year 4 could be interpreted as a more rapid transition to a new sequence type in the plasma. The possible mechanisms and the consequences of this observation are discussed in sections 4.4.4 and 4.4.5.

The relative frequencies of sequences differing in length in the V4 region were also estimated by high-resolution gel electrophoresis of amplified non-diluted cDNAs (containing 100-200 copies of target sequences) and DNAs (containing approximately 100-220 target molecules), providing a more comprehensive analysis of the variants

present in the samples than would be obtained by sequencing alone. The relative numbers of *group b* and *group c* sequences on the polyacrylamide gel were then quantified by scanning densitometry (carried out by Drs. P. Simmonds and A.J. Leigh Brown, see Figure 4.9). The relative proportions agree closely with those determined by sequence analysis. For example, at year 4 (1988), DNA samples contained 74 to 75% *group b* sequences while the cDNA samples contained only 42 to 48% (Simmonds *et al.*, 1991). The corresponding numbers of *group b* sequences are 10 out 11 and 3 of 11 in these two samples. Similarly, the DNA samples at year 5 (1989) contained 84 to 85% *group c* sequences by densitometry, compared with 10 of 11 by sequence analysis, while the cDNA population was uniformly *group c* by both methods (Simmonds *et al.*, 1991).

4.3.11 Linkage and the recombination of the V3 Loop and the V4 region of gp120 Examination of the sequential sequences obtained in the V3 and V4 regions of patient 82 reveals some other interesting phenomena. Despite the continuous sequence change of the overall plasma viral RNA population with time, some variants with identical V3 loop persist and actually increase their frequency over the whole population. This is especially the case for the E1 and F genotype sequences (see Table 4.4). While the frequency of F genotype sequences kept more or less constant over the study period (13.3% in year 3, 9.1% in year 4 and 7.7% in year 7), the proportion of the E1 genotype sequences increased from 6.7% in year 3 to 33.3% in year 6, and in year 7 it began to dominate the plasma viral population (76.9%). Furthermore, the reappearance of some early sequences in the later samples was also

Figure 4.9. Length analysis of the V4 hypervariable region to confirm the existence of population differences in the *in vivo* proviral DNA and viral RNA populations. Lane: a, negative human DNA amplified with primers (w, v and x, u) spanning the V4 region (see Figure 2.1); b and d, PCR product from V4 region of proviral DNA at years 4 and 5, respectively; c and e, PCR product from viral RNA in the corresponding plasma samples. Expected sizes of *group b* and *group c* sequences are indicated as B and C, respectively (Simmonds *et al.*, 1991).


a b c d e
C
B

observed in the V3 region. After being completely undetectable for over two years (years 5 and 6) in the plasma viral population, a sequence of class F, which is quite common at year 3, re-emerged at year 7 (see Table 4.4). Similarly, the reappearance of the early *group b* sequences at years 6 and 7 was also observed after it was completely replaced by *group c* sequences at year 5 (see Figure 4.6, Molecules Z82r86, Z82r95, Z82r97 and Z82r101).

To address whether this pattern is the consequence of linkage or recombination, the V3 and V4 viral RNA sequences obtained from the same PCR amplified cDNA molecules were studied. A total of 68 full-length V3-C2-V4 viral RNA sequences were obtained (approximately 400bp in length). The genotypes for both V3 and V4 regions of the same molecule are presented in Figure 4.10. There is no fixed relationship between V3 and V4 genotypes. V3 loop sequences of the E1 type contain both *group b* sequences, Z82r05 for example, or *group c* sequences at V4 (Z82r71, see Figure 4.10). Conversely, sequences with the same V4 genotype could have completely different V3 genotype sequences. For example, *group c* sequences at V4 could be linked to V3 loop sequences either of E1 (Z82r28), or E2 (Z82r29), or D5 (Z82r26) genotypes. In view of the substantial difference between both V4 *group b* and V4 *group c* sequences and between V3D and V3E sequences, the lack of a complete association between them implies that recombination must occur repeatedly. The hypothesis is supported by the recent discovery that viral recombination indeed occurred frequently among variants found in a single HIV-1 positive individual (Howell *et al.*, 1991). New recombinants may have some selective advantages over pre-existing ones and, if so, will reach high frequency in plasma population until they

Figure 4.10 Genotype comparison of the V3 and V4 regions from the same molecules. Sequence genotypes for the V3 and V4 region were determined according to previously defined sequence types (see sections 4.3.6 and 4.3.8). Year 0 = 1984; year 3 = 1987; year 4 = 1988; year 5 = 1989; year 6 = 1990 and year 7 = 1991.

Year	Sequences	genotypes	
		V3	V4
3	Z82r01	C1	B1
	Z82r02	C2	B1
	Z82r03	C3	B1
	Z82r04	B	B1
	Z82r05	E1	B1
	Z82r06	C2	B1
	Z82r07	F	B1
	Z82r08	F	B1
	Z82r09	C2	B1
4	Z82r10	D1	B1
	Z82r11	F	C1
	Z82r12	E2	C1
	Z82r13	D3	C1
	Z82r15	E2	C1
	Z82r16	D2	C1
	Z82r17	D2	C1
	Z82r18	E3	C1
	Z82r19	D2	C1
	Z82r20	D2	B1
	Z82r21	D2	B1
5	Z82r22	E2	C1
	Z82r24	D4	C1
	Z82r25	E4	C1
	Z82r26	D5	C1
	Z82r27	E5	C1
	Z82r28	E1	C1
	Z82r29	E2	C1
	Z82r30	E2	C1
	Z82r31	E2	C1
	Z82r32	E2	C1
	Z82r33	E2	C1
	Z82r34	E2	C1
	Z82r35	E2	C1
	Z82r37	E2	C1
	Z82r38	E2	C1
	Z82r39	E2	C1
	Z82r40	E2	C1
	Z82r41	E2	C1
	Z82r42	D4	C1
	Z82r43	E2	C1
	Z82r44	E5	C1
	Z82r45	E2	C1
	Z82r46	E2	C1
6	Z82r70	D8	C9
	Z82r71	E1	C2
	Z82r72	E6	C7
	Z82r76	E7	C7
	Z82r77	D8	C7
	Z82r78	E8	C1
	Z82r79	D6	C5
	Z82r80	E1	B3
	Z82r81	D7	C1
	Z82r82	E1	C8
	Z82r83	D7	C5
	Z82r84	D7	C7
	Z82r85	E1	C7
	Z82r86	E1	B1
7	Z82r91	D7	C1
	Z82r92	E1	C6
	Z82r94	E1	C2
	Z82r95	E1	B1
	Z82r96	E1	C2
	Z82r97	E1	B1
	Z82r98	E1	C2
	Z82r100	D8	C1
	Z82r101	E1	B1
	Z82r102	F	C1
	Z82r103	E1	C1

too encounter an immune response. Within the relatively homogenous plasma viral population observed at year 5 (1989) (see Table 4.2 and 4.3) this is likely to be the case. Another consequence of the variable association between these two hypervariable regions is that the frequencies of sequences varied independently from each other. Combination of the V3 and V4 sequences will show a higher rate a turnover than that of the different sequences considered separately. Also, this will lead to even greater differences between the plasma viral and lymphocyte-associated proviral population at a given time point.

4.4 Discussion

Analysis of the plasma viral RNA sequences obtained from serial samples of a single haemophiliac patient (patient 82) has revealed tremendous sequence diversity in both the V3 and V4 hypervariable regions. This rapid sequence evolution consists not only of emergence and disappearance of distinct viral variants, but also of frequency changes of these variants over time, indicating persistent changes in genetic composition of the plasma viral population. This finding was anticipated by our previous cross-sectional studies of sequence evolution in a cohort of haemophiliac patients infected from a common source (Balfe *et al.*, 1990). Similar results were also obtained from the study of the V3 sequence variation of six children infected from a single plasma donation by Wolfs *et al.*, (Wolfs *et al.*, 1990). However, one of the most striking findings of the analysis of patient 82 is the observation that the spectrum of variation develops from a homogenous population of V3 sequences at seroconversion. Examination of V3 and V4 sequences obtained from the same

molecule reveals that there is very restricted number of sequence linkages between these two regions. Lack of a complete association between the V3 and V4 region suggests that a recombination event must occur. Because of the functional importance of the V3 region, and also because the V3 is a major target for both the antibody and cytotoxic T-cell recognition, it is reasonable to assume that the evolutionary process of the V3 region may be different from that of the V4 region, with more selective constraints on the V3 than on the V4 region (see section 4.3.4). Significant differences between the frequencies of sequence variants in plasma viral and lymphocyte-associated proviral populations within the same sample were also observed for both V3 and V4 regions, suggesting that at any one time point, the predominant plasma virus variants may be antigenically different from sequences encoded by proviral DNAs present in PBMCs. The implications of the above findings for our understanding of the biology of HIV are discussed below.

4.4.1 Low Frequency of Inactivating Substitutions In this study, approximately 35,000 bases of viral RNA were directly sequenced from patient materials and no inactivating substitutions were found. This finding is consistent with our previous studies of sequence variation of proviral DNA sequences present in PBMCs amongst members of Edinburgh haemophiliac cohort where the calculated frequency of inactivating substitutions is approximately 1 in every 20 kilobases (Balfe *et al.*, 1990). Low frequency of inactivating substitutions has also been reported for those infected by vertical transmission (Wolinsky *et al.*, 1992). Therefore, it is likely that the inactivating substitution is extremely low in the *in vivo* situation, and that defective

virus variants may not play a major role in the pathogenesis of HIV. This result is somewhat in contrast to that described by Meyerhans *et al.* (Meyerhans *et al.*, 1989), who observed that the frequency of detective mutations was about 10-fold higher than that found in our experiments. The discrepancy between the results could be due to the fact that different parts of HIV-1 genome were assessed in each study or due to the different technical procedures applied to obtain nucleotide sequences. Their sequences were obtained from clones of isolates propagated in *in vitro* culture whereas our sequences were obtained directly from PCR products. It has been known for some time that *in vitro* culturing virus will result in reducing the variability of virus variants (Meyerhan *et al.*, 1989; Kusumi *et al.*, 1992), and that the cloning of PCR products into plasmid vectors could amplify possible errors generated during the process of PCR amplification (Simmonds *et al.*, 1990*a*). Thus the high frequency of defective virus variants observed by Meyerhans *et al.*, could partly be an artifact generated during *in vitro* manipulation.

4.4.2 Effective Immune Response at Early Stage of HIV Infection In this work, clear evidence has been presented for the rapid decline of plasma viral load during the early stage of infection (see section 4.3.1). This result was anticipated by the studies of viral and proviral loads in homosexual men during the period of primary infection, where the virus load in both plasma and PBMCs population has been shown to decrease rapidly and spontaneously (Daar *et al.*, 1991; Clark *et al.*, 1991). Although it is possible that other factors maybe involved in the decline of viral load, it is likely that the fast reduction in virus quantity *in vivo* reflects the development of an effective

immune reaction in the host. At the present stage, it is difficult to determine the immunological factors that govern viral replication during the early period of HIV infection. Recent studies, however, favour the hypothesis that an effective response from cytotoxic T-cells may play major role in restricting viral replication and transmission. In an investigation of the cell-mediated immune response to HIV-1 in seronegative homosexual men with recent sexual exposure to HIV-1, cytotoxic T lymphocyte responses were detected in the absence of a humoral immune response (Clerici *et al.*, 1991 and 1992). The sharp decline of plasma viraemia was also observed before the appearance of neutralising antibodies (Ariyishi, St Mary's Hospital Medical School, personal communication).

4.4.2 *In Vivo* Sequence Evolution of the V3 and V4 regions in Patient 82 It has been shown in this work that there is a high degree of sequence variation in the V3 and V4 hypervariable regions of gp120 in patient 82. In the V3 region, virtually all the sequence variation is due to frequent nucleotide substitutions whereas in the V4 region, deletion and/or insertion events are also involved. However, the substitutions do not seem to accumulate steadily with time. This is reflected by the presence of an extremely homogenous plasma viral population observed at year 5 (1989) for both the V3 and V4 regions (see section 4.3.4).

Sequence evolution in the V3 region is a complex process. There is no simple relationship between successive plasma viral RNA sequences, nor is there clear replacement of one V3 genotype with another. The evolutionary tree of the V3 region obtained repeatedly by different phylogenetic methods (see section 4.2.4) shows a

striking and consistent evolutionary pattern. All the later sequences can be explained as progeny of the sequences found at seroconversion. Three years after seroconversion, a major phylogenetic division into two distinct lineages has been observed and both of which persisted throughout the course of the infection. The rapid and continuous evolution of each lineage results in persistent changes of the genotype distribution and therefore of the genetic composition of the plasma viral population. It has been argued for some time that the rapid sequence change in the V3 region is a consequence of a continual process of immune escape. Several studies, including this work, have indeed shown that high rates of amino acid replacement are precisely located in those areas that can be recognized by the immune system (see section 4.3.6.; and Holmes *et al.*, 1992). Furthermore, frequent sequence changes lead to rapid alterations in the relative distribution of different genotypes, suggesting that the frequency-dependent element may well be involved in the evolutionary process of plasma viral population. In this aspect, the process of sequence changes could be partially explained by a simple host-parasite model; the higher the frequency a viral variant reaches in the plasma population, the higher probability of its recognition and clearance by the immune system. From the data obtained in this work, it can be clearly seen that in each year investigated it is indeed the most frequent genotype that appears to be cleared away (shows the greatest reduction in frequency). Consequently, variants found at low frequency will have a greater selective advantage and will rise in frequency until they too are countered with an effective immune response. The frequency dependent nature of viral evolution is made more complicated by the possibilities that; 1) the selection process may not be constant in the long term

because of the loss of normal functions of B and T cells that are associated with prolonged HIV infection, and 2) the qualitative changes in virus phenotype may result in the emergence of virus variants that are more virulent and cytopathic for CD4+ T cells (Cheng-Mayer *et al.*, 1988; Tersmette *et al.*, 1989*a, b*).

The pattern of sequence variation in the V4 region is slightly different. Apart from frequent amino acid replacement through time, length variation is also observed in this region. Sequence change in patient 82 consists of a series of replacements of one particular sequence type with another over the first 5 years of infection. However, in years 6 (1990) and 7 (1991) after infection, multiple types of V4 sequences coexist in the plasma viral population. Furthermore, some viral variants found in early infection reappeared in the plasma viral population probably due to the loss of normal memory functions of immune system. Supporting this hypothesis is the observation that CD45RO+ memory cells appears to be preferentially infected *in vivo* by both HIV and SIV (Schnittman *et al.*, 1989 and 1990; Willerford *et al.*, 1990). Sequence analysis of the V4 region reveals that the succeeding V4 sequence type was not necessarily directly derived from the previous sequence; for example, *group c* succeeded *group b* at year 4 (1988) to year 5 (1989), yet *group b* probably was not the immediate ancestor of any of the *group c* sequences which reflected by the less evolutionary distance between *group a* and *group c* than that of between *group a* and *group b* (see section 4.3.4). The discontinuous evolutionary pattern observed in plasma is likely to be a consequence of the evolutionary process of virus variants in different solid tissues. CD4+ cells in solid tissues have long been shown to be infected frequently (Fauci, 1992; Meltzer *et al.*, 1990) and recent discoveries have

indicated that the distinct pattern of V3 and V4 variants are present in different organs of HIV-1 positive patients (Ball *et al.*, Regional Virus Laboratory, East Birmingham Hospital, Birmingham, personal communication; Dr. P. Simmonds, personal communication; Epstein *et al.*, 1991; Steuler *et al.*, 1992). Contribution to the plasma viraemia by these distinct V4 sequences at different times and/or at different levels may therefore partially explain the replacement of one V4 sequence type with another as well as the existence of the discontinuous evolution of this region.

4.4.3 Constraints on the V3 Sequences Having analyzed the nucleotide and amino acid sequences of the V3 region, a unique pattern of sequence variation has been observed. Apart from its general hypervariable nature shown above, high conservation of some amino acid segments were identified within the V3 region, implicating their functional importance. This is especially true for the amino acid segments CTRPNNNTRK, IGPG, and IRQAHC which constitute the frame-structure of the V3 loop. Indeed, several biological functions have been proposed for the V3 region from the *in vitro* study of biological clones with mutations in the V3 loop. It has been suggested that the V3 region is implicated in virus-cell fusion, possibly mediated by protease cleavage near the tip of the loop (Clements *et al.*, 1991). The region encompassing the V3 loop has also been shown to be the major determinant of T-cell and macrophage tropism (Cheng-Mayer *et al.*, 1990*a,b*; Hwang *et al.*, 1991). Certain amino acid mutations, particularly in the conserved regions of the V3 loop, could generate a virus that has either lost or substantially reduced its infectivity in the *in vitro* assays (Bolognesi, 1990).

This functionally critical region can, at the same time, be recognized by both neutralizing antibodies and cytotoxic T cells (Rusche *et al.*, 1988; Palker *et al.*, 1988; Cease *et al.*, 1987; Takahashi *et al.*, 1988). Although the amino acid sites between conserved segments are highly variable and have been postulated as one of the major means of viral escape from immune clearance, detailed sequence analysis has revealed that only a limited number of amino acid are allowed to change at these sites. Phylogenetic analysis has demonstrated that identical amino acid changes have occurred in independent lineages, suggesting extensive convergent evolution is implicated in the process of V3 sequence variation (this thesis and Holmes *et al.*, 1992). In this aspect, it is likely that two major elements are governing the evolution process of the V3 sequences. One is the hypervariable element which will facilitate viruses to escape once they are recognized by the immune system. The other is the restricting element which would allow the sequence changes as long as these changes are functionally viable for further virus infection.

4.4.4 Long-term Persistence of Seroconversion Sequences Analysis of sequence variation in the plasma viral population and comparison with that of lymphocyte-associated proviral populations revealed the existence of differences in the frequencies of different sequence types of virus present in these two populations (see section 4.3.10). While substantial sequence evolution has taken place in the plasma viral RNA population, several proviral DNA sequences obtained several years (up to 3 years in patient 82) after primary infection were still identical to those detected at seroconversion. This finding suggests that these proviral DNA sequences correspond

to the seroconversion-type sequences may not replicate to any significant extent during the intervening years. Although there is a sharp decline in the number of CD4+ cells around seroconversion (Clark *et al.*, 1991; Daar *et al.*, 1991), a fraction of these infected cells which are somehow resistant to the cytopathic effect of HIV can survive and continue to circulate in the blood stream. It is possible that these cells are a subset of memory cells which appear to be preferentially infected *in vivo* by both HIV and SIV (Schnittman *et al.*, 1989; Willerford *et al.*, 1990). Consistent with their function in antigenic recall, these memory cells can have essentially unlimited life span (Simmonds *et al.*, 1991). Thus, differentiated memory cells will persist *in vivo* for quite a long time bearing within them unchanged seroconversion proviral sequences. The persistence of *group b* DNA sequences in patient 82, when almost all RNA sequences were of group c type, might also have been the consequence of long-term persistence of cells non-productively infected in 1987 (see section 4.2.9).

There are several possible explanations for the proposed long-term survival of infected T-cells (Simmonds *et al.*, 1991). Firstly, proviral sequences in those PBMCs that survive infection may contain inactivating mutations that prevent subsequent virus replication. High frequencies of defective proviral sequences have been reported to exist *in vivo* (Meyerhans *et al.*, 1989). However, using the limiting dilution PCR method that can eliminate *in vitro* copying errors during amplification, an extremely low rate of inactivating substitutions has been observed (see section 4.3.3). Furthermore, it has been demonstrated that a high proportion of proviral sequences present in PBMCs can be activated *in vitro* to give replication-competent viruses (Brinchmann *et al.*, 1991). Therefore, defective viruses probably contribute little to

persistent infection of lymphocytes. An alternative explanation for the failure of HIV to kill the cell it infects is that virus which exists at the early stage of infection may be less virulent and cytopathic for T cells. This notion is supported by the emergence of HIV variants that are more T-cell cytopathic *in vitro* as the disease progresses. The phenotype of these variants was classified as 'rapid/high' (Tersmette *et al.*, 1989*a, b*). Cell tropism studies of HIV-1 from a number of investigators has also suggested that Macrophage-and-T-cell (MT) tropic isolates are the predominant HIV variants detected early after infection of humans, whereas T-cell-tropic isolates became more prevalent as the disease progresses (Hwang *et al.*, 1991). These phenotypic changes have been shown to correlate with HIV envelope sequence change, especially in the V3 region (Hwang *et al.*, 1991). The isolates from primary HIV infection which are defined as MT-tropic V3 loop sequences are all similar to each other, and to the 'global consensus' V3 loop sequence identified from 245 isolates from North American (Hwang *et al.*, 1991). In contrast, T-cell-tropic isolates appear to be characterized by the V3 loop sequences that are dissimilar to this 'global consensus'. The sequence analysis in this work also fits above hypothesis. The seroconversion V3 sequences obtained from patient 82 are identical to the 'global consensus' while sequences found at later stages are all clearly different from the 'global consensus'. Taken together, the above observations are consistent with the idea that the envelope-gene evolution, particularly in the V3 region, may reflect the selection of T-cell-tropic and T-cell cytopathic variants with nonconsensus V3 region sequences during the later stage of HIV-1 induced disease. Less virulent and less T-cell cytopathic viral variants at seroconversion may be one of the reasons for long-term persistence of seroconversion

sequences in the proviral DNA population. Thirdly, it could also be because of the silent nature of the host cells. As has been know for some time, viral replication requires an "activated" T cell environment and the presence of various inducible host transcription factors. Quiescent T cells, which make up a large proportion of the T cell pool *in vivo*, appear deficient in these factors (Zack *et al.*, 1990*a, b*; Bukrinsky *et al.*, 1991; Stevenson *et al.*, 1990; Varmus *et al.*, 1977). Thus, as one of the latent forms (see section 1.1.9), seroconversion proviral sequences can also persist in the body for quite a long period.

4.4.5 Origin of Plasma Viral Population Differences between plasma viral and lymphocyte-associated proviral populations were observed in this work. The more rapid sequence turnover in the plasma population than in PBMCs also suggests that the new virus variants may appear and become dominant firstly in the plasma population. Consistent with this hypothesis, the sequence variants found in the earliest plasma sample (seroconversion sequence) were not seen subsequently in the plasma, but were relatively abundant 3 years later as provirus in the PBMCs of patient 82 (Simmonds *et al.*, 1991). With disease progression, the proportion of infected PBMCs will increase and the proviral population will begin to reflect the more recent plasma sample. Consequently, the proportion of PBMCs bearing the seroconversion sequence type will become smaller and smaller.

There are several possible explanations for the original source of the plasma viral population. First of all, it could be from a subset of transcriptionally active CD4+ lymphocytes. It has been shown that plasma of both symptomatic and

asymptomatic individuals is infectious, and thus infection of PBMCs may be a self-sustaining process. However, the only small proportion of infected CD4+ cells in the blood and an even smaller proportion of CD4+ actively expressing HIV mRNA in the course of infection can hardly explain the plasma abundances of cell-free circulating virus *in vivo* and the high titre of plasma virus infectivity *in vitro* (Simmonds *et al.*, 1990*a*; Ho *et al.*, 1989; Coombs *et al.*, 1989). Although, in most cases, blood monocytes have been documented at normal levels even when patients were in late-disease stage (Poli *et al.*, 1985), examination of tissue macrophages in HIV-1 infected patients suggests a completely different picture. In certain body tissues, such as those of the central nervous system, lymph nodes, or lung, the frequency of HIV-1 infected cells may be 10,000- to 100,000-fold higher than that in blood. In each of these tissue, the predominant cell type infected is not the CD4+ T cells, but rather the macrophages (Meltzer *et al.*, 1990). Recent studies have suggested that the predominant HIV-1 isolates obtained early after infection are more macrophage-tropic, whereas T-cell-tropic isolates became more prevalent as disease progress (Ross *et al.*, 1992; Schuitemaker *et al.*, 1991 and 1992). Taken together, it may suggest that, in the early infection, macrophage may serve as reservoir for virus and also as a vehicle for virus dissemination. Large proportions of infected macrophages in certain tissues may therefore account for the main source of plasma virus through the course of infection, although other origins such as dendritic cells in the blood circulation cannot be ruled out.

CHAPTER 5

SEQUENCE VARIABILITY OF PLASMA VIRAL RNA AND LYMPHOCYTE-ASSOCIATED PROVIRAL DNA DURING THE PRIMARY INFECTION OF HIV-1

5.1 Summary

An investigation was undertaken of the levels of plasma viraemia and the sequence variation of HIV-1 during the primary stage of infection. Five plasma samples were obtained from five independently infected individuals prior to, or immediately after seroconversion. Four PBMCs samples were also collected 3-6 months after seroconversion from 4 members of the Edinburgh haemophiliac cohort who were infected from a single common batch of HIV-1 contaminated factor VIII (see section 1.3, Chapter 1).

High levels of plasma viraemia were observed in all samples studied ranging from 10^5 - 10^8 virus particles per ml of plasma, indicating an explosive replication of viral population at very early stage of HIV infection. Five years after infection, however, there was a substantial drop in viral load in plasma, suggesting an effective and rapid immune response limiting viral replication *in vivo*.

Surprisingly, there is no sequence variation in the V3 region within any of the samples studied. This finding contrasts strongly to the situation seen in the follow-up samples where substantial sequence variation was observed. The closely linked V4 region is also lacking in sequence variation in all but one cases investigated. At the same stage, however, sequence variation *is* present in the p17 region of *gag*, a genomically distant region from the V3 and V4 regions. Sequence analysis of the V3 sequences from these pre- and post-seroconversion samples has revealed their substantial similarity to the 'global consensus' V3 loop sequence identified from 245 North American isolates (LaRosa *et al.*, 1990), and maybe to the sequences carried by some monocyte-tropic virus variants (Fouchier *et al.*, 1992; Chesebro *et al.*, 1992;

De Jong *et al.*, 1992*a, b*; Westervelt *et al.*, 1991 and 1992). In addition, 3 out of 4 members of the Edinburgh haemophiliac cohort have identical V3 and V4 proviral sequences and the forth patient has only one nucleotide difference in the V3 region from the others. The uniformity of the V3 and V4 regions compared to the heterogeneity in the p17 region strongly implies that 1) viral infection is not initiated by a single viral particle, 2) the homogeneity of the V3 and V4 regions is not because of the random outgrowth of certain viral genotypes, and 3) strong selective pressure has been imposed on the envelope gene in the initial phase of viral infection.

As the V3 region contains one of the major targets for both B and T cells, identifying those sequences which are selected for in the early stage of viral infection will be important not only in terms of documenting the viral sequence changes which can overcome the specific immune response, but also in providing sequence information for the design and development of an effective vaccine, especially if such sequence homogeneity in *env* found in blood stream can also be extended to viruses found in different organs during the same stage of HIV infection.

5.2 Introduction

Primary infection with HIV-1 is most commonly manifested as an acute illness characterised by fever, myalgia, rash, gastro-intestinal symptoms, and occasionally neurological manifestations (Ho, *et al.*, 1985). The period from the onset of acute illness to seroconversion can range from eight days to three months (Ho *et al.*, 1985), with virus detectable in the semen, cerebrospinal fluid, peripheral

blood mononuclear cells (PBMCs), and plasma before the development of an antibody response (Ho *et al.*, 1985; Goudsmit *et al.*, 1986; Tindall *et al.*, 1992). During this period high levels of p24 antigen has been detected at levels comparable to or even higher than those detected in patients in the later stage of HIV infection (Goudsmit *et al.*, 1986; McRae *et al.*, 1991). High levels of viraemia have been found in both plasma and lymphocyte populations (Clark *et al.*, 1991; Daar *et al.*, 1991; Jurriaans *et al.*, 1992) implying that HIV-1 is capable of explosive replication *in vivo* in the very early stage of infection. Following seroconversion, however, a rapid decline of cell-free virus in plasma and of lymphocyte-associated provirus in PBMCs was observed in an interval of 9 to 39 days and 6 to 34 days respectively (Daar *et al.*, 1991; Jurriaans *et al.*, 1992). This indicates the presence of an highly effective and rapid anti-HIV immune response at this time. Furthermore, the titers of infectious virus in plasma and PBMCs samples from seroconversion patients, measured by the end-point-dilution method, are very high and some are even substantially higher than those from some symtomatic patients (Daar *et al.*, 1991; Clark *et al.*, 1991), suggesting that the majority of the virus variants present during the primary stage of infection are infection-competent (Dr. David Ho, VIII International Conference on AIDS, Amsterdam, The Netherlands, 1992). The patients therefore possess an enormous amounts of infectious viruses at the primary infection and will be highly infectious, particularly as anti-HIV antibodies are undetectable by current antibody-based immunosorbent assay (Daar *et al.*, 1991; Leigh Brown, personal communication).

More interestingly, phenotypic studies of those virus variants obtained from

patients during the primary stage of HIV infection reveal that nearly 85% of them are non-syncytium-inducing (NSI) variants and preferentially infect monocyte-lineage cells (Roos *et al.*, 1992). Such findings are consistent with the study of Cichutek *et al.*, who demonstrated that seroconversion viruses could not be isolated in T-cell lines such as Molt-4/8, MT-4 and CEM, but were isolated when purified primary human macrophage were used (Cichutek *et al.*, 1991). Recently, Schuitemaker *et al.*, have shown that the progression towards AIDS is associated with a shift of the viral population from monocyte- to T-cell-tropic variants (Schuitemaker *et al.*, 1991 and 1992). Such a finding also supports the notion that monocyte-tropic NSI variants are predominant in the primary stage of HIV infection (Tersmette *et al.*, 1989*a, b*; Ross *et al.*, 1992). Therefore, low-level yet persistent infection of monocytes may therefore have greater selective advantage and survive the potent anti-HIV-1 immune response during the early period of infection.

There is currently little information on the sequences of virus variants present in the early stage of HIV infection. Such data will be of great value as it will provide information in the study of sequential sequence variation *in vivo*, and will enable to trace those mutations at the molecular level which may determine the changes in viral tropism for particular cell types and escape from the immune response. At the same time, such information will also provide valuable information for the design and development of vaccine.

In this part of the work, attention has been drawn exclusively to the study of sequence variability of plasma viral and lymphocyte-associated proviral populations during the stage of primary infection, aiming to provide some insight into the process

affecting the viral population during this period. In addition to the seroconversion sample obtained from patient 82 (see section 4.1), another 4 plasma samples were collected from 4 independently infected patients (p74, p84, Sc1 and Sc2) prior to or immediately after seroconversion. Patients 82, 74 and 84 represent haemophiliacs in Edinburgh whereas patients Sc1 and Sc2 are two cases of heterosexual transmission from Edinburgh (1988) and from Dundee (1991), respectively. Four PBMC samples were also collected 3-6 months after seroconversion from 4 members (p28, p77, p79 and p84) of the Edinburgh haemophiliac cohort who were infected from a single common batch of HIV-1 contaminated factor VIII (Ludlam *et al.*, 1985). Sequence analysis is largely concentrated on the V3 and V4 regions of the *env* gene, and part of the *p17* region of the *gag* gene.

5.2 Results

Seroconversion plasma samples from 5 independently infected patients and three following-up plasma samples out of five patients (p82, p74 and p84) were obtained from Drs. Roy Robertson (Edinburgh City Hospital), Christopher Ludlam (Edinburgh Royal Infirmary), G.E.D. Urquhart and A.J. France (both Department of Medical Microbiology, University of Dundee) via Dr. Peter Simmonds (Department of Medical Microbiology, University of Edinburgh). No patient, except patient 74 is currently on, or has previously received any anti-viral treatment. Four PBMCs samples from 4 members of Edinburgh haemophiliac cohort were also obtained. Plasma viral RNA titre, quantified by previously developed RNA based PCR, the p24

antigen level in plasma and the serological status of anti-HIV-1 antibody of the 5 independently infected patients are present in Table 5.1. Apart from patient 82 whose antibody to HIV proteins was weakly positive when sample was collected, the other four patients were all negative for anti-HIV antibodies.

5.2.1 High Levels of Plasma Viraemia During the Primary Infection of HIV-1 Comparison of plasma viral loads at seroconversion and five years after has revealed a striking difference (Table 5.1). The viral loads are 3-5 orders of magnitude higher at seroconversion than after 5 year of infection. For example, there were approximately 10^8 and 10^5 virus particles per ml of plasma at seroconversion in patient 82 and 84, respectively. However, five years after infection, plasma viral loads were dropped to 10^3 in patient 82 and was not detectable in patient 84 (<100, see Table 5.1). High levels of plasma viraemia observed in seroconversion plasma samples suggests that HIV-1 could replicate rapidly and efficiently in a short period of time. Substantial decline of viral loads after 5 years infection, however, may indicate the potent anti-HIV-1 activity of the immune system at early stage of infection. The persistence of viral production in plasma after several years of infection, however, may equally suggest that the latent period observed clinically is not necessary reflected by non-replication of viruses *in vivo*.

5.2.2 Plasma Viral RNA Sequences of *env* and *gag* Genes During the Primary Infection Plasma viral RNA were extracted and detected by an RNA based PCR technique developed previously in this work (see Chapter 3). Single cDNA

Table 5.1 Viral RNA titre, p24 antigen level and serological antibody status in plasma samples of five individuals collected during the stage of primary infection. [a] virion titre in the plasma were quantified by RNA-based PCR method developed in this work. [b] p24 antigen level and [c] anti-HIV-1 antibodies results were kindly provided by Dr. Peter Simmonds and Mrs. Selma Rebus (Department of Medical Microbiology, University of Edinburgh). NA, not applicable.

Patients	Virion per ml plasma by PCR[a] at seroconversion / 5 years after	p24 Ag (pg/ml)[b]	Antibody to HIV[c]
82	1.0×10^8 / 8.5×10^3	0	weakly positive
74	1.0×10^7 / 3.49×10^3	50	negative
84	1.0×10^5 / $< 1.0 \times 10^2$	20	negative
Sc1	1.0×10^4 / NA	15	negative
Sc2	1.0×10^8 / NA	500	negative

molecules were obtained by limiting dilution prior to the double PCR amplification (see section 2.2.7). Plasma viral RNA sequences were obtained by directly sequencing the PCR amplified products without any further *in vitro* manipulation (see section 2.2.10). From those seroconversion plasma samples, 12 V3 loop sequences were determined from patient 82, 22 from patient 84, 40 from patient 74, 26 from S1c, and 29 from S2c. These nucleotide sequences are presented in Figure 5.1, together with their respective deduced amino acid sequences lined underneath. The V4 plasma viral RNA sequences from these 5 samples have also been obtained and they are presented in Figure 5.2. From patient 82, 9 V4 sequences were determined, whereas from patients 84, 74, S1c and S2c, 22, 21, 26, and 11 sequences were obtained respectively. Part of *p17* region of the *gag* gene, when the samples were available, were also determined. From patient 74, total of 15 sequences were obtained while from patients 84, Sc1 and Sc2, 8, 19 and 18 sequences were obtained respectively. These *gag* sequences are presented in Figure 5.3.

5.2.3 Proviral DNA Sequences of *env* and *gag* genes from 4 Members of the Edinburgh Haemophiliac Cohort Proviral DNA were extracted from 4 members of the Edinburgh haemophiliac cohort as described (see section 2.2.3). Detection and sequencing of proviral DNA molecules from these PBMCs samples were carried out as described previously (see section 2.2.10). The V3 and V4 sequences were obtained from the same HIV-1 DNA molecules. From patient 28, 11 V3 and V4 sequences were obtained while from patients 77, 79 and 84, 11, 9 and 11 V3 and V4 sequence were obtained respectively. For the p17 region of the *gag* gene, 9

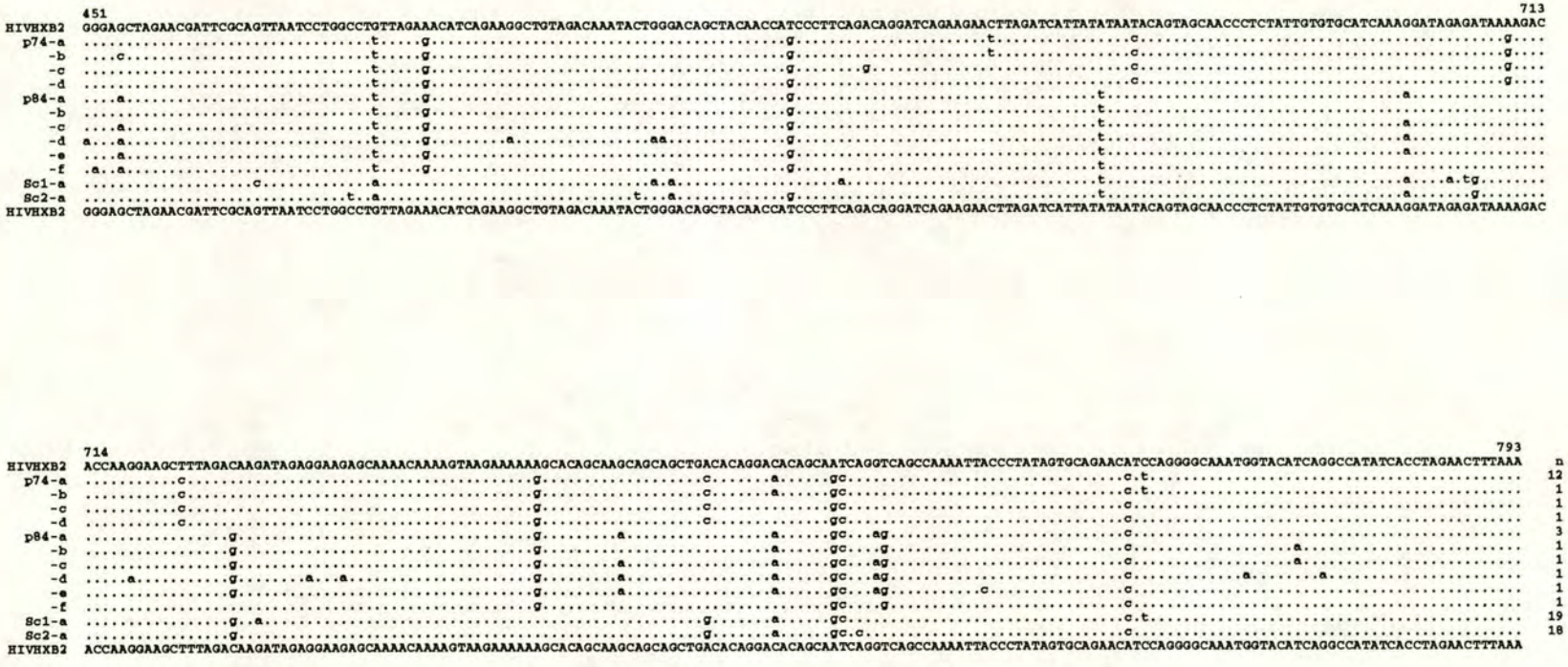
Figure 5.1 Nucleotide and deduced amino acid sequences of the V3 loop obtained at or immediately after seroconversion. Amino Acid positions are numbered according to HIV_{HXB2}. V3 loop sequence from HIV_{MN} is presented according to the HUMAN RETROVIRUSES and AIDS 1991 database by Los Alamos National Laboratory, Los Alamos, New Mexico 8754, USA. The 'global V3 loop' consensus sequence identified from 245 North America isolates is also presented (LaRosa *et al.*, 1990).

Patient	V3 Loop Sequences 296 330	No. of Sequences obtained	No. of Sequence Variants
82	TGTACAAGACCCAACAACAATACAAGAAAAAGTATACATATAGGACCAGGAAGAGCATTTTATACAACAGGAGAAATAATAGGAGATATAAGACAAGCACATTGT .C..T..R..P..N..N..N..T..R..K..S..I..H..I..G..P..G..R..A..F..Y..T..T..G..E..I..I..G..D..I..R..Q..A..H..C.	12	1
74	TGTACAAGACCCAGCAACAATACAAGAAAAAGTATACATATGGGACCGGGGAGAGCATTTTATGCAACAGGAGAAATAATAGGAGATATAAGACAAGCACATTGT .C..T..R..P..S..N..N..T..R..K..S..I..H..M..G..P..G..R..A..F..Y..A..T..G..E..I..I..G..D..I..R..Q..A..H..C.	40	1
84	TGTACAAGACCCAGCAACAATACAAGAAGAAGTATATCAATAGGACCGGGGAGAGCATTTTATGCAACAGGAGAAATAATAGGAGATATAAGACAAGCACATTGT .C..T..R..P..S..N..N..T..R..R..S..I..S..I..G..P..G..R..A..F..Y..A..T..G..E..I..I..G..D..I..R..Q..A..H..C.	22	1
Sc1	TGTACAAGACCCAACAACAATACAAGAAAAAGTATACATATAGGACCAGGCAGAGCATTTTATACAACAGGACGAATAATAGGAAATATAAGACAAGCACATTGT .C..T..R..P..N..N..N..T..R..K..S..I..H..I..G..P..G..R..A..F..Y..T..T..G..R..I..I..G..N..I..R..Q..A..H..C.	26	1
Sc2	TGTACAAGACCCAACAACAATACAAGAAAAGGTATACATATAGGACCAGGCAGAGCATTTTATACAACAGGAGAAATAATAGGAGATATAAGACAAGCATATTGT .C..T..R..P..N..N..N..T..R..K..G..I..H..I..G..P..G..R..A..F..Y..T..T..G..E..I..I..G..D..I..R..Q..A..Y..C.	29	1
HIV_{MN}	.C..T..R..P..N..Y..N..K..R..K..R..I..H..I..G..P..G..R..A..F..Y..T..T..K..N..I..I..G..T..I..R..Q..A..H..C.	1	1
LaRosa.con	.C..T..R..P..N..N..N..T..R..K..S..I..H..I..G..P..G..R..A..F..Y..T..T..G..E..I..I..G..D..I..R..Q..A..H..C.	1	1

Figure 5.2 Nucleotide and deduced amino acid sequences of the V4 region obtained at or immediately after seroconversion. Amino acid positions are numbered according to HIV_{HXB2}.

Patient	V4 Sequences 389 423	No. of sequences obtained	No. of sequence variants
82	TTTAATAGTACTTGGAATTCAACACAACTTAATAGTACTTGGAATTCAACACAACTTAATAGTGCTGGGAATAATACTGAAGAAAATATCACACTCCCATGTAGA .F..N..S..T..W..N..S..T..Q..L..N..S..T..W..N..S..T..Q..L..N..S..A..G..N..N..T..E..E..N..I..T..L..P..C..R.	1	4
	TTTAATAGTACTTGGAATTCAACACAATTTAATAGTACTTGGAATTCAACACAACTTAATAGTGCTCGGACTGAAGAAAATCTCACACTCCCATGTAGA .F..N..S..T..W..N..S..T..Q..F..N..S..T..W..N..S..T..Q..L..N..S..A..R..T..E..E..N..L..T..L..P..C..R.	1	
	TTTAATAGTACTTGGAATTCAACACAACTTAATAGTACTTGGAATAATACTGAAGAAAATATCACACTCCCATGTAGA .F..N..S..T..W..N..S..T..Q..L..N..S..T..W..N..N..T..E..E..N..I..T..L..P..C..R.	6	
	TTTAATAGTACTTGGAATTCAACACAATTTAATAGTACTTGGAATTCAACACAACTTAATAGTGAAGAAAATATCACACTCCCATGTAGA .F..N..S..T..W..N..S..I..Q..F..N..S..T..W..N..S..T..Q..L..N..S..E..E..N..I..T..L..P..C..R.	1	
74	TTTAATAGTACTTGGAATAATAATGATACTAGTACTTGGAATGAGACTGGAAAGTCAGATAACATCACACTCCCATGCAGA .F..N..S..T..W..N..N..N..D..T..S..T..W..N..E..T..G..K..S..D..N..I..T..L..P..C..R.	21	1
84	TTTAATAGTACTTGGAATGATACTACAGGGTCAAATACTACAGGGTCAAATAACACTGAAACTATCACACTCCCATGCAGA .F..N..S..T..W..N..D..T..T..G..S..N..T..T..G..S..N..N..T..E..T..I..T..L..P..C..R.	22	1
Sc1	TTTAATAGTATTTGGAAGGTTAATAGTACTTGGAATGGTACTGGAGGATCAAATAACACGGAAGGAAAGGACACAATCACACTCCCATGCAGA .F..N..S..I..W..K..V..N..S..T..W..N..G..T..G..G..S..N..N..T..E..G..K..D..T..I..T..L..P..C..R.	26	1
Sc2	TTTAATAGTACTTGGAATGGTAATGGTACTTGGGATGTTACTGGAGGGTCAAATAACACTGAAGGAAATGACACAATCACACTTCCATGCAGA .F..N..S..T..W..N..G..N..G..T..W..D..V..T..G..G..S..N..N..T..E..G..N..D..T..I..T..L..P..C..R.	11	1

Figure 5.3. Viral RNA Sequences of the p17 region of *gag* gene in patients 74, 84, Sc1 and Sc2 during the primary infection. Sequence of HIV_{HXB2} is obtained from HUMAN RETROVIRUSES and AIDS 1991 database by Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA and is used for sequence comparison. Nucleotide sequence positions are numbered according to HIV_{HXB2}. Only nucleotides that differ from HIV_{HXB2} are shown. Dots denote identical nucleotides to HIV_{HXB2}. 'n' is the number of each variant sequence obtained.


451
713
HIVHXB2 GGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGAC
p74-a
-b
-c
-d
p84-a
-b
-c
-d
-e
-f
Sc1-a
Sc2-a
HIVHXB2 GGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGAC
714
793
HIVHXB2 ACCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAAGCACAGCAAGCAGCAGCTGACACAGGACACAGCAATCAGGTCAGCCAAAATTACCCTATAGTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAA n
p74-a 12
-b 1
-c 1
-d 1
p84-a 3
-b 1
-c 1
-d 1
-e 1
-f 1
Sc1-a 19
Sc2-a 18
HIVHXB2 ACCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAAGCACAGCAAGCAGCAGCTGACACAGGACACAGCAATCAGGTCAGCCAAAATTACCCTATAGTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAA

sequences were obtained from patient 28, 9 from patient 77, 6 from patient 79 and 8 from patient 84. The deduced amino acid sequences from the V3 and V4 regions of *env* and nucleotide sequences of p17 region of *gag* gene are presented in Figures 5.4, and 5.5, respectively.

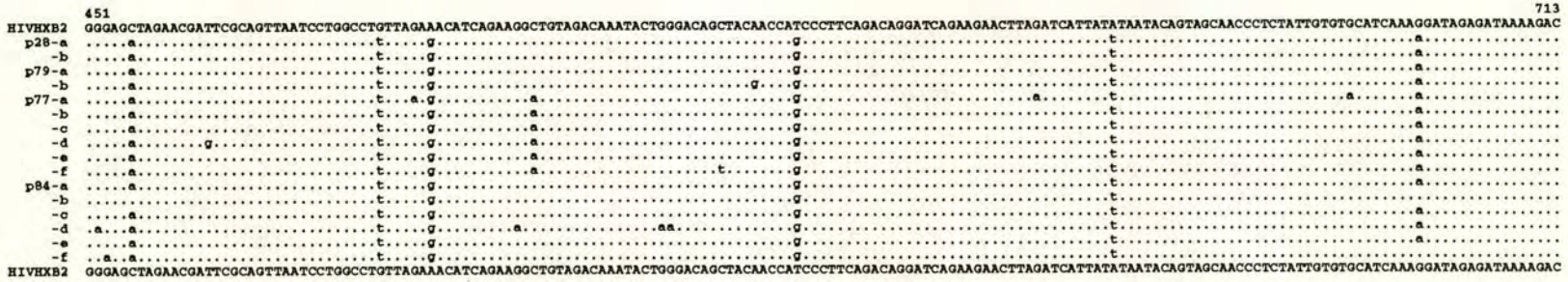
5.2.4 Plasma Viral RNA Sequences from Follow-up Samples For the purpose of sequence comparison, three follow-up plasma samples were obtained from patients 82, 74 and 84. From these samples, 23 V3 loop sequences were obtained from patients 82, 6 from patient 74 and 15 from patient 84. These sequences were aligned and the final nucleotide sequence alignment is shown in Figure 5.6 (nucleotide sequences from patient 82 are presented in Figure 4.1, see section 4.2.3).

5.2.5 Lack of Sequence Variation in the V3 Hypervariable Region During the Primary Infection Examination of nucleotide and amino acid sequences presented in Figures 5.1 and 5.4 reveals a striking feature of sequence variation during the first stage of HIV infection. No sequence variation in the V3 hypervariable region in any of the pre- and post-seroconversion samples was observed. Three out of four members of the Edinburgh haemophiliac cohort have identical V3 proviral sequences and the forth patient (p77) has only one residue difference from the others (see Figure 5.4). This finding contrasts strongly to the situation observed in the later on samples where a high degree of sequence variation has been observed (Figure 5.6). In one patient (p84) whose pre-seroconversion plasma sample and post-seroconversion PBMC sample are both available, the sequences obtained from the V3 region are

Figure 5.4 Deduced amino acid sequences of the V3 and V4 regions from four members of the Edinburgh haemophiliac cohort. These amino acid sequences are direct translation of proviral DNA sequences obtained from 4 samples 3-6 months after seroconversion. Gaps between the V3 and V4 regions are indicated by dashes. Only residues that differ from p28 are shown. Dots denote identical residues to p28 sequence. 'n' is the number of sequences obtained from each patient. N.V., number of sequence variant observed.

Patient	V3 loop	V4 region	n	N.V.
p28	CTRPSNNTRRSISIGPGRAFYATGEIIGDIRQAHC	-----FNSTWNDTTGSNTTGSNNTETITLPCR	11	1
p79	.....	-----.....	9	1
p77	.....P.....	-----.....	11	1
p84	.....	-----.....	11	1

Figure 5.5 Proviral DNA Sequences of the p17 region of *gag* gene in 4 members of the Edinburgh haemophiliac patients (p28, p79, p77 and p84) during the primary infection. Sequence of HIV_{HXB2} is obtained from HUMAN RETROVIRUSES and AIDS 1991 database by Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA and is used for sequence comparison. Nucleotide sequence positions are numbered according to HIV_{HXB2}. Only nucleotides that differ from HIV_{HXB2} are shown. Dots denote identical nucleotides to HIV_{HXB2}. 'n' is the number of each variant sequence obtained.


451
713
HIVHXB2 GGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGAC
p28-a
-b
p79-a
-b
p77-a
-b
-c
-d
-e
-f
p84-a
-b
-c
-d
-e
-f
HIVHXB2 GGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGAC

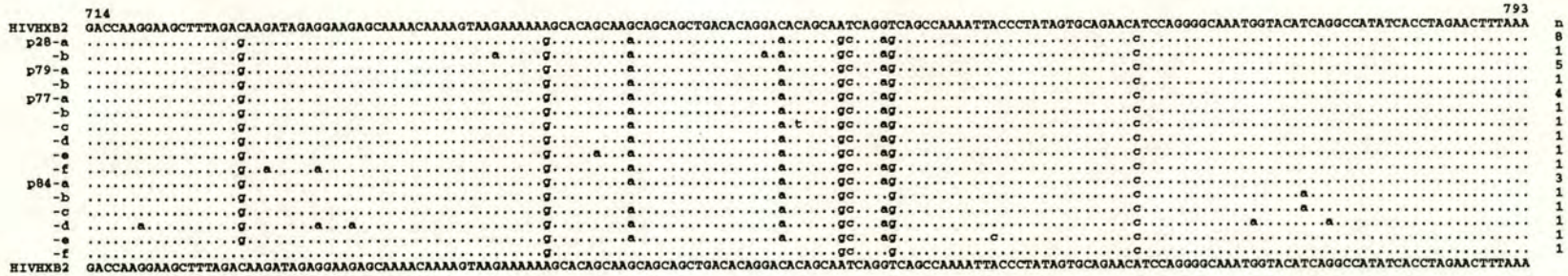

714
793
HIVHXB2 GACCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAAGCACAGCAAGCAGCAGCTGACACAGGACACAGCAATCAGGTCAGCCAAAATTACCCTATAGTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAA n
p28-a 8
-b 1
p79-a 5
-b 1
p77-a 4
-b 1
-c 1
-d 1
-e 1
-f 1
p84-a 3
-b 1
-c 1
-d 1
-e 1
-f 1
HIVHXB2 GACCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAAGCACAGCAAGCAGCAGCTGACACAGGACACAGCAATCAGGTCAGCCAAAATTACCCTATAGTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAA

Figure 5.6 Nucleotide sequences of the V3 loop obtained from follow-up sampled of patients 74 and 84. Nucleotide sequences of the V3 region from patient 82 after 5 years infection are presented in Figure 4.1 (a) (Chapter 4). Consensus sequence are generated by most frequent nucleotide at each position. Nucleotide positions are numbered according to the HIV_{HXB2} sequence. Each sequence represents a single viral RNA molecule obtained directly from plasma. Only nucleotides that differ from the consensus are shown. Dots denote identical nucleotides to the consensus sequences.

```
Sequences   888                                                                947
74v38916    .......... .......... ........a. .......... .......... ..........
74v38917    .......... .......... .......... .......... .......... ..........
74v38918    .......... .......... .......... .......... .......... ..........
74v38920    .......... .......... .......... .......... .......... ..........
74v38921    .......... .......... ........a. .......... .......... ...t......
74v38923    .......... .......... .......... .......... .......... ..........
Consensus   TGTACAAGAC CCAACAACAA TACAAGAAGA GGTATACATA TAGGACCAGG GAGAGCATTT

            948                                                    992
74v38916    .......... ..g....... .......... .......... .....
74v38917    .......... .......... .......... .......... .....
74v38918    .......... .......... .......... .......... .....
74v38920    .......... .......... .......... .......... .....
74v38921    .......... ..g....... .......... .......... .....
74v38923    .......... .......... .......... .......... .....
Consensus   TATGCAACAG GAAACATAAT AGGAGATATA AGACAAGCAC ATTGT
```

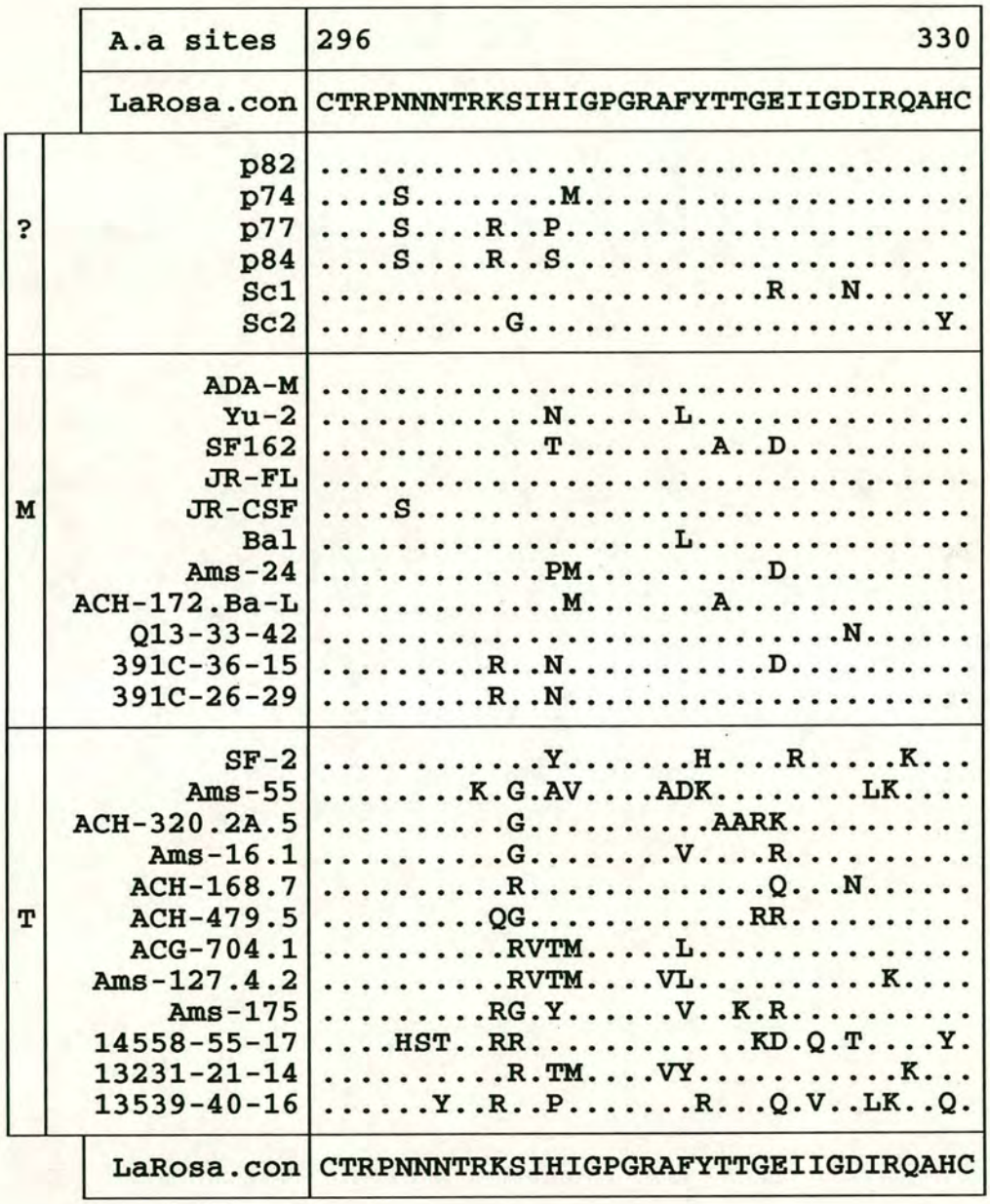
```
Sequences   888                                                                947
84v31       .......... .......... .......... ...t...... .......... ..........
84v35       .......... .......... .......... .......... .......... ..........
84v37       .......... .......... .......... .......... .......... ..........
84v310      .......... .......... .......... .......... .......... ..........
84v312      .......... ..aa...... .......... ..t....... .......... ..........
84v319      .......... .......... .......... .......... .......... ..........
84v323      .......... .......... .......... .......... .......... ..........
84v325      .......... .......... .......... .......... .......... ..........
84v3r02     .......... .......... .......... .......... .......... ..........
84v3r03     .......... .......... .......... .......... .......... ..........
84v3r04     .......... .......... .......... .......... .......... ..........
84v3r06     .......... .......... .......... .......... .......... ..........
84v3r08     .......... ..a....... .......... ..t.....c. .......... ..........
84v3r10     .......... ..aa...... .......... ..t....... .......... ..........
84v3r12     .......... .......... .......... .......... .......... ..........
Consensus   TGTACAAGAC CCGGCAACAA TACAAGAAAA AGGATATCAA TAGGACCGGG GAGAGCATTT

            948                                                    992
84v31       .......... g......... .......... .......... .....
84v35       .......... .......... .......... .......... .....
84v37       .......... .......... .......... .......... .....
84v310      .......... .......... .......... .......... .....
84v312      ta.......g g.g.t..... .......... ...c...... .....
84v319      .......... .......... .......... .......... .....
84v323      .......... .......... .......... .......... .....
84v325      .......... .......... .......... .......... .....
84v3r02     .......... .......... .......... .......... .....
84v3r03     ...a...... g......... .......... .......... .....
84v3r04     .......... .......... .......... .......... .....
84v3r06     .......... .......... .......... .......... .....
84v3r08     ta.......g g.g....... .......... ...c...... .....
84v3r10     ta.a.....g g.g.t..... .......... ...c...... .....
84v3r12     ta.......... .......... .......... .......... .....
Consensus   ATTGCAACAA AACAAATAAT AGGAGATATA AGAAAAGCAC ATTGT
```

identical between these two populations, suggesting sequences present in the post-seroconversion DNA sample, to a large extent, may represent the situation of those pre-seroconversion RNA samples.

Moreover, substantial similarity has been found in the V3 loop sequence between seroconversion amino acid sequences and 'global consensus' V3 loop sequence identified from 245 North American isolates (LaRosa *et al.*, 1990). In one case (p82), the amino acid sequence is identical to the 'global consensus' V3 loop sequence. Furthermore, if these seroconversion V3 sequences are compared with those carried by virus variants with distinct phenotypes, it appears that the seroconversion sequences may be more similar to those sequences carried by monocyte-tropic virus variants (see Figure 5.7) than to those possessed by T-cell-tropic virus variants, suggesting virus variants present during the primary stage of infection may have some common biological characteristics.

Examination of V4 sequences obtained from these pre- and post-seroconversion samples reveals similar results. There was no sequence variation in 4 (patients 84, 74, S1c, and S2c) out of 5 plasma samples studied while in one case (patient 82), 6 out of 9 sequences were the same (see Figure 5.2). Sequence heterogeneity observed in the V4 region in patient 82 could be due to the fact the this sample was collected slightly after seroconversion (antibody was weakly positive from HIV-1, see Table 5.1). Unexpectedly, proviruses from the 4 members of the Edinburgh haemophiliac cohort bear identical V4 sequences, suggesting: 1) that these 4 haemophiliac patients were indeed infected from the same source and 2) that the homogeneity of the envelope gene during seroconversion can not be due to the

Figure 5.7 Comparison of V3 loop sequences obtained from pre- and post-seroconversion samples with those carried by distinct phenotypic virus variants identified in the *in vitro* culture (Fouchier *et al.*, 1992; Chesebro *et al.*, 1992). M, V3 loop sequences from macrophage tropic variants. T, sequences from T cell-tropic variants. ?, indicates the uncertainty about virus phenotype bearing these seroconversion V3 loop sequences. LaRosa.con., V3 loop consensus sequence of 245 North American isolates (LaRosa *et al.*, 1990). Only residues that differ from LaRosa.con are shown. Dots denote identical residues to LaRosa.con. Amino acid positions are numbered according HIV_{HXB2}.


A.a sites 296 330
LaRosa.con CTRPNNNTRKSIHIGPGRAFYTTGEIIGDIRQAHC
? p82 .....
p74 ....S.....M.....
p77 ....S....R..P.....
p84 ....S....R..S.....
Sc1 .....R...N.....
Sc2 .....G.....Y.
M ADA-M .....
Yu-2 .....N.....L.....
SF162 .....T.....A..D.....
JR-FL .....
JR-CSF ....S.....
Bal .....L.....
Ams-24 .....PM.....D.....
ACH-172.Ba-L .....M.....A.....
Q13-33-42 .....N.....
391C-36-15 .....R..N.....D.....
391C-26-29 .....R..N.....
T SF-2 .....Y.....H....R.....K...
Ams-55 .....K.G.AV....ADK.....LK....
ACH-320.2A.5 .....G.....AARK.....
Ams-16.1 .....G.....V....R.....
ACH-168.7 .....R.....Q...N.....
ACH-479.5 .....QG.....RR.....
ACG-704.1 .....RVTM.....L.....
Ams-127.4.2 .....RVTM....VL.....K....
Ams-175 .....RG.Y.....V..K.R.....
14558-55-17 ....HST..RR.....KD.Q.T....Y.
13231-21-14 .....R.TM....VY.....K...
13539-40-16 .....Y..R..P.....R...Q.V..LK..Q.
LaRosa.con CTRPNNNTRKSIHIGPGRAFYTTGEIIGDIRQAHC

random growth of one particular virus variants.

Analyzing sequences obtained from the *gag* p17 region, which are presented in Figures 5.3 and 5.5, has revealed another striking pattern. In contrast to the situation observed in the V3 and V4 regions where no sequence changes have been encountered, sequence variation in the *gag* gene was frequently found. Sequences obtained from *gag* p17 region, including those from plasma viral RNAs and PBMC-associated proviral DNAs, have shown a comparable degree of variation in 5 out of 7 patients studied, with exception from patients Sc1 and Sc2 where no sequence variation has so far been observed. The number of sequences obtained and the numbers of sequence variants from the V3, V4 and p17 regions are summarized in Figure 5.8. In most the cases (p74, p28, p77, p79 and p84) while there are only one sequence variant in the V3 and V4 regions, more than one sequence variants were observed in the *gag* p17 region. How this finding contributes to our understanding of HIV infection and transmission will be discussed below.

5.3 Discussion

5.3.1 Rapid Changes of Viral Loads During the Primary Stage of HIV-1 Infection One of the major features associated with primary HIV-1 infection is the rapid changes of virus loads in both plasma and PBMCs populations. Irrespective of the route of transmission, high levels of viraemia in both plasma and PBMCs appears to be a constant phenomenon during the primary stage of HIV-1 infection. The quantification results presented in this work has once again confirmed the

Figure 5.8 Comparison of *env* and *gag* sequence variation during the primary and secondary infection. [a] RNA, viral RNA sequences obtained from pre-seroconversion plasma samples; DNA, proviral DNA sequences obtained from PBMC samples collected 3-6 months after seroconversion. [b] Pr, samples were collected at pre-seroconversion; Po, post-seroconversion and Fo, following-up samples collected 5 years after seroconversion. N.A., not applicable. N.D. not done.

patient	Sample[a]	Stage[b]	V3 Region Total No. sequences / variants	$p17_{gag}$ Region Total No. sequences / variants
p74	RNA	Pr	40 / 1	15 / 4
p82	RNA	Pr	12 / 1	N.A.
p84	RNA	Pr	11 / 1	N.A.
Sc1	RNA	Pr	26 / 1	19 / 1
Sc2	RNA	Pr	29 / 1	18 / 1
p28	DNA	Po	11 / 1	9 / 2
p79	DNA	Po	9 / 1	6 / 2
p77	DNA	Po	11 / 1	9 / 6
p84	DNA	Po	11 / 1	8 / 6
p82	RNA	Fo	23 / 5	N.D
p74	RNA	Fo	6 / 3	N.D
p84	RNA	Fo	15 / 6	N.D

extremely high levels of plasma viraemia during the first stage of HIV-1 infection. Dramatic increase in virus loads to the level which is only comparable to, or even higher than that found in patients with AIDS or AIDS-related complex, suggests HIV-1's capability of explosive replication within a short period of time. These results lead to the suggestion that the majority of the virus variants present during the primary stage of infection are infection-competent (Dr. David Ho, VIII International Conference on AIDS, Amsterdam, the Netherlands, 1992). These findings, as suggested by Drs. David Ho and Leigh Brown, imply that there is a window period during the primary stage of HIV-1 infection when the patient has tremendous amount of infectious virus although anti-HIV antibodies are not detectable by current antibody-based immunosorbent assay. The pre-seroconversion patients therefore are highly likely to be one of the major sources for the transmission of HIV-1 (Daar *et al.*, 1991; Leigh Brown, personal communication).

However, the extremely high level of virus loads during the acute infection is temporary and transient. Not long after the infected host developed specific immune response to HIV-1 proteins, marked drop of virus loads were observed in both plasma and PMBCs populations, suggesting that immune response during the period is effective and efficient (Daar *et al.*, 1991; Clark *et al.*, 1991; Jurriaan *et al.*, 1992). Although in general, it is believed that both humoral and cellular immunity have both played roles in limiting the viral replication during the primary stage of HIV-1 infection, recent reports, however, suggest that anti-HIV-1 specific T-cell response may play the major role. Clerici *et al.*, have demonstrated the appearance of anti-HIV-1 specific T-cell response prior to the emergence of serum anti-HIV-1

antibodies (Clerici *et al.*, 1991 and 1992). Ariyishi *et al.*, have also shown that the sharp decline of plasma viraemia happened before the appearance of neutralising antibodies (Ariyishi, St. Mary's Hospital Medical School, personal communication). At present stage, there is not enough data to elucidate the relationship between the appearance of anti-HIV-1 humoral and cellular immunity and their relative roles in restricting viral replication, yet it will be a important step if the precise immune mechanisms that are responsible from the potent anti-HIV-1 activity can be defined. A better understanding of these processes will help us to exploit the beneficial anti-HIV-1 immune responses to develop more effective treatment strategies for AIDS.

5.3.2 Genetic Variability of HIV-1 Genome During the Primary Infection In this work clear evidence has been presented for the lack of sequence variation in the V3 and V4 regions of *env* in both plasma viral and PBMC-associated proviral populations during the primary stage of HIV-1 infection. This finding contrasts strongly to what has been observed for the same regions in the later stages of infection, where enormous sequence diversity have been frequently found (see Chapter 4). In the *gag* gene, however, sequence variation has been observed in the most cases studied including those sequences directly obtained from both plasma and peripheral blood mononuclear cells.

There are several possible explanations from the homogeneity of the envelope gene during the primary stage of HIV-1 infection. First of all, viral infection could be initiated by a single virus. The first and successful infection of certain cell type with a single virus variant will subsequently result in homogeneous population

because of the rapid explosive replication of the virus variant. Low transmission rate of HIV-1 infection through sexual contact (Clumeck *et al.*, 1989) may indeed support the notion that the HIV-1 infection is initiated only by a limited number of virus variants present in the inoculum. However, the sequence variation observed in the *gag* gene at the same period can hardly be reconciled with this hypothesis. Different viral sequences found in the *gag* gene suggests that the viral population is a heterogenous one during the early stage of primary infection. Secondly, the genetic homogeneity in the envelope gene could be due to the random outgrowth of a certain virus variant. This hypothesis can be explained by population genetic theory. As the effective population number (N_e) of an expanding population is much lower than the census number (N), the enormous increase in population size would itself be expected to cause a substantial reduction in nucleotide diversity on a simple neutral model (Crow and Kimura, 1970). The enormous increase in viral population size that occurs during the period immediately after infection is therefore itself be expected to cause substantial reduction in nucleotide diversity (Dr. A. J. Leigh Brown personal communication). In other words, unless a substantial numbers of virus variants were sampled, then a homogeneous population would be expected to be observed because of sampling errors associated with the fast expanding viral population. However, considerable sequence similarity among 4 members of the Edinburgh haemophiliac cohort infected from a single common source tends to rules out the second possibility. The sequence similarity among the seroconversion sequences and to the 'global consensus' V3 loop sequences and perhaps to the sequences carried by those monocyte-tropic virus variants favour the third theory, which implies that there

appears some kind of directional selection 1) genetically for sequences which resemble the 'global consensus' V3 loop sequence and 2) biologically for those virus variants with monocyte-tropic capability (see Figure 5.7). If this is the case, then the sequence variation observed in the *gag* gene will not have any effect on the initiation and initial stages of viral infection and the *env* gene will be one of the major targets on which the action of multiple selective forces is imposed. This notion is supported by the recent discovery that 1) majority of virus variants isolated during the primary stage of HIV-1 infection are indeed monocyte-tropic as indicated by their failure to grow on T-cell lines and their non-syncytium-inducing (NSI) phenotype (Ross *et al.*, 1992; Cichutek *et al.*, 1991; Daar *et al.*, 1991) and 2) seroconversion V3 region sequences, particularly the V3 loop sequences, obtained worldwide so far, are very similar (Drs. Peter Simmonds and Edward Holmes, personal communication). According to this hypothesis, successful initiation of HIV-1 infection is therefore dependent on qualitative criteria which certain V3 loop genotypes viruses must posses like viral tropism for certain cell types and less virulent and cytopathic effect on the infected host cells. If this is really the case, it will suggest that: 1) whatever virus variants are present in the inoculum, those that are monocyte-tropic will be selected during the primary infection; 2) the monocyte-tropic HIV-1 isolates are important for viral persistence during the early stage of infection and for dissemination of HIV-1 to compartments outside the peripheral blood; 3) the virus variants existing in tissue macrophages, which are more frequently infected than those present in the peripheral blood (Koening *et al.*, 1986; Gendelman *et al.*, 1989; Meltzer *et al.*, 1990*a, b, c*), are the progeny of variants selected during the early stage of infection and arising later,

probably serve as one of the main reservoirs contributing to the plasma viraemia; and 4) that in terms of early diagnosis, treatment and vaccine development, the monocyte-tropic virus variants and their host cells should be the target for future therapeutic strategies.

CHAPTER 6

GENERAL DISCUSSION

6.1. LEVEL OF CIRCULATING HIV RNA IN PLASMA IS A GOOD VIROLOGICAL MARKER FOR HIV INFECTION AND EFFICACY OF ANTI-VIRAL THERAPY

6.2 QUANTITATIVE AND QUALITATIVE VARIABILITY OF HIV DURING THE COURSE OF INFECTION *IN VIVO*

6.3 ORIGIN OF PLASMA VIRAEMIA

6.1. LEVEL OF CIRCULATING HIV RNA IN PLASMA IS A GOOD VIROLOGICAL MARKER FOR HIV INFECTION AND EFFICACY OF ANTI-VIRAL THERAPY

The natural history of HIV infection is manifested by a number of changes in virological and immunological markers, together with the gradual onset and resolution of clinical symptoms (Phillips, 1992; Fauci, 1988). Among these markers, the levels of cell-associated virus (Oka *et al.*, 1990; Schnittman *et al.*, 1991; Simmonds *et al.*, 1990*a*), p24 antigenaemia (Allain *et al.*, 1986; Goudsmit *et al.*, 1986), CD4 lymphocytes (Goedert *et al.*, 1987; Laga *et al.*, 1989), viral antibody (Allain *et al.*, 1986; Nishanian *et al.*, 1990), and infectious plasma viraemia (Ho *et al.*, 1989; Coombs *et al.*, 1989; Simmonds *et al.*, 1991) are the most commonly used references for monitoring the course of viral infection as well as the outcome of clinical treatment. In this work, a new quantitative technique has been developed which can be used to detect, quantify and sequence HIV RNA genomes directly from plasma and serum of HIV-1 positive individuals. Thereby, more direct virological information can be readily obtained. Using this technique, the amount of HIV RNA present in plasma and serum was quantified. On average, HIV RNA is more abundant in the plasma of patients with more advanced disease compared with asymptomatic (see section 3.2.4). A sharp decrease in plasma viral load has been observed after patients received anti-viral treatment (AZT) (data not shown). Similar results have also been obtained from other laboratories (Ho *et al.*, 1989; Coombs *et al.*, 1989).

Compared with the other virological and immunological markers, the amount of circulating HIV RNA measured by the RNA-PCR based techniques reflect more

accurately the status of those replicative viruses *in vivo*. p24 antigenaemia is detectable at certain stages of HIV infection, and is probably one of the most commonly used markers for monitoring the course of HIV infection (Allain *et al.*, 1986). However, viral protein p24 can be partially or completely complexed with antibodies (Lange *et al.*, 1987; Ujhelyi *et al.*, 1987), and thus undetectable by conventional antigen assay. The levels of p24 antigen detected, in this sense, do not reflect directly the rate of viral replication but rather the joint outcome of p24 and anti-p24 antibody production. The levels of CD4+ cell (expressed as either an absolute number, or a percentage of lymphocytes, or a ratio of CD4+ to CD8+ T cells) have been shown to be one of the best single predictor of the progression to AIDS and has been recommended as a marker for anti-viral treatment (Phillips, 1992; Schnittman *et al.*, 1989). However, by the time when there is a obvious reduction of CD4+ cells, in most cases, patients have already entered the stage where any anti-viral treatment may not have any effective results. Ho *et al.*, and Coombs *et al.*, have measured the infectious viral titre by *in vitro* culturing of primary lymphocytes with plasma, and have found that around 3,000 infectious particles per ml plasma in symptomatic patients, and around 30 per ml plasma in asymptomatic (Ho *et al.*, 1989; Coombs *et al.*, 1989). Although the use of *in vitro* culture technique to quantify infectious HIV-1 is very specific, it is very time consuming and tiresome, and therefore, may not be suitable for quantifying large numbers of samples.

The RNA-PCR based quantitative technique developed in this work is rapid, efficient and independent of the presence of virus-antibody complexes (see Chapter 3), and can be used before the serological response developed (see Chapter 5). Using

this technique, transient high levels of plasma viraemia were observed during the primary infection, suggesting the explosive replication of incoming viruses during a very short period of time. High and intermediate levels of plasma viraemia were also observed in some patients during their asymptomatic stage, indicating viral replication continues throughout the course of infection. Virologically, there is no latent stage to correspond to the clinical latent period observed in almost all the HIV-1 positive patients. In addition, this technique also provides a means of detection and quantitation of viruses present in other body fluids such as semen and saliva (Tindall *et al.,* 1992; Barr *et al.,* 1992), as well as in blood products such as factor VIII and factor IX concentrates. With direct sequencing of the PCR products, the virus sequences present in the original batch of blood products can be determined, and that knowledge will be extremely valuable for the understanding of the transmission and variation of HIV. However, one problem facing the RNA-PCR based quantitative technique, as facing any other PCR related techniques, is the possible cross-contamination from either other samples and/or recombinant plasmid containing HIV sequences. Extra care must always be taken, and proper negative and positive controls have to be included in the assay.

6.2 QUANTITATIVE AND QUALITATIVE VARIABILITY OF HIV DURING THE COURSE OF INFECTION *IN VIVO* Sequential studies of viruses present in both plasma viral and lymphocyte-associated proviral populations reveal an extra-ordinary degree of variability in both virus loads and sequences. When viruses firstly enter the human body, they will always undergo a rapid and explosive replication in

a very short period of time, which is reflected by a transient high level of virus loads in both plasma and lymphocytes (Daar *et al.*, 1991; Clark *et al.*, 1991). Because of the lack of specific immune recognition of any specific viral antigens during this period, the rapid expansion in viral population suggests that there is a strong selection for the most rapidly replicating viral variant, with the consequent loss of variation due to either selection and/or genetic linkage.

Sequence analysis of plasma viral RNA sequences obtained prior to seroconversion and proviral DNA sequences present in PBMCs 3-6 months after seroconversion have shown the existence of homogenous viral and proviral sequences in the envelope gene, while noticeable sequence variation was found in the comparable *gag* sequences, indicating that selection early in the infection acts most strongly on the envelope gene. Furthermore, the observation that the V3 loop sequences in patient 82 at seroconversion was identical to the 'global consensus' V3 loop sequence and those of other patients very similar probably suggests some biological significance of the seroconversion virus variants. Although the detailed studies of biological phenotypes of seroconversion viruses has yet to come, the available information at present related to these viruses suggests that there is a strong selection for viruses with certain biological features. One of these features may be the preferential infection of macrophages or monocyte-derived cells. This notion is strongly supported by recent experiments showing that virus isolates obtained from seroconversion samples have failed to grow in T cell lines, but replicate well in PBMCs and macrophage cell lines in the *in vitro* culture (Cichutek *et al.*, 1991; Ross *et al.*, 1992). With the progression of infection, a shift of viral population from

macrophage-tropic to T-cell-tropic has been observed which further confirmed the macrophage-tropic feature of seroconversion viruses (Tersmette *et al.*, 1989; Hwang *et al.*, 1991; Schuitemaker *et al.*, 1992). Macrophages, therefore, may serve as a major reservoir for virus during the early stage of infection and also as a vehicle for virus dissemination to different organs. Virus variants homing at different organs may have different genetic and biological features at the later stages of infection (Cheng-Mayer *et al.*, 1990*a*; Ball *et al.*, Regional Virus Laboratory, East Birmingham Hospital, Birmingham, personal communication). HIV variants present in brain have been shown to differ biologically from those exist in the blood circulation at the later stage of infection (Cheng-Mayer *et al.*, 1990*a*; Hwang *et al.*, 1991). A study of viral phenotype of SIV_{mac} from different organs suggested that the viruses localized in spleen, lymph nodes and plasma are lymphocyte-tropic while those found in brain and lungs are macrophage-tropic (Sharma *et al.*, 1992). However, the genetic and biological characteristics of HIV variants present in different organs during the primary stage of infection are not yet known. If the homogeneity in both genetic and biological features of seroconversion virus variants found in blood circulation can be extended to those viruses localized in different organs during the same period, this will indeed bring some encouraging news for the design and development of an effective vaccine against AIDS.

The high proportion of infected CD4+ cells and high levels of plasma viraemia observed during the primary stage of infection will sooner or later bring about a response of the immune system. A sharp reduction in the levels of both plasma viraemia and HIV infected CD4+ cell during the seroconversion period have

been observed, suggesting a strong effect of the immune response. Although the immune response during this period has not been extensively studied, recent reports suggest that anti-HIV-1 T-cell cytotoxicity may play an important role (Clerici *et al.*, 1991 and 1992). Sequence variability of HIV after seroconversion is very different from what observed before seroconversion. A highly heterogenous viral population is generated and sequence variation in the envelope gene has been observed within a couple of weeks (Pang *et al.*, 1992). In the work presented here, a major diversification of V3 sequences from patient 82 has been observed in plasma within 3 years from seroconversion. It has frequently been argued that the rapid sequence changes in the envelope gene, particularly in the third hypervariable region, are the consequence of continual process of immune escape. Albert *et al.*, have shown the emergence of virus variants *in vivo* that were resistant to neutralization by autologous sera (Albert *et al.*, 1990). 'Escape' mutants of HIV-1, due to selection for a point mutation in the V3 loop, have also been selected by neutralizing antibody *in vitro* (McKeating *et al.*, 1989). Other studies, including this work, have indeed shown the high rates of amino acid replacements precisely located in the areas that are the targets for both the serological and cytotoxic T cells recognition (Phillips *et al.*, 1991).

Rapid evolution of the viral population during this period occurs also by changes in the relative abundance of different sequences (see chapter 4), suggesting that the frequency-dependent element is involved in the evolutionary process of viral population. Thus the process of sequence change is in accordance with antigenic variation can be explained by a simple host-parasite model of frequency-dependence;

the higher the frequency a viral variant reaches in the population, the higher probability of its recognition and clearance by the immune system. From the data obtained in this work, it can be clearly seen that in each year it is indeed the most frequent virus variants that show the greatest reduction in frequency. Consequently, variants found at low frequency will have a greater selective advantage and will rise in frequency until they too are countered with an effective immune response. Finally, the assessment of antigenic diversity in the V3 loop with consequent escape from immune recognition is complicated by the fact that mutations outside principal neutralization domain (PND) can also confer resistance to immune clearance (McKeating *et al.,* 1989), suggesting that the PND is an conformational epitope. Evidence that the V3 region is an conformational epitope was demonstrated by successful isolation of several neutralizing antibodies which can recognize and bind to the V3 region irrespective of amino acid sequences variation (Steimer *et al.,* 1991; Sattentau *et al.,* 1991; Ho *et al.,* 1991).

The immune defect caused by HIV infection is progressive and irreversible. Virtually all the HIV-1 positive individuals will finally enter the symptomatic stage although some patients may stay on in asymptomatic stage for years (Fauci, 1988). The turning point is surely determined by multiple factors, including at least the effectiveness of host immune system and the degree of pathogenicity of invading viruses. Based on the current available information of interaction between the virus and the immune system, Nowak *et al.,* have proposed a mathematical model which suggests the existence of an antigen diversity threshold below which the immune system is able to control viral population growth but above which the virus

population will induce the collapse of the immune system (Nowak *et al.,* 1990, 1991*a, b*). Quantitative analysis of virus loads in both plasma viral and lymphocyte-associated proviral populations indeed demonstrated the existence of more and viruses in symptomatic than asymptomatic patients (see Chapter 3). Viral sequence of the envelope region of gp120 shows greater variability during the symptomatic than asymptomatic stage (see chapter 4). Antigenic diversity overrides the effectiveness of immune responses and could therefore be one of the explanations of the onset of AIDS. However, the process of viral evolution during this stage becomes more complicated by the fact: 1) that the selection process may not be constant in the long term because of the loss of normal functions of B and T cells during prolonged HIV infection and, 2) that the qualitative changes of viral phenotype may result in the emergence of virus variants that are more virulent and cytopathic for CD4+ T cells (Hwang *et al.,* 1991; Schuitemaker *et al.,* 1992). The sequence and antigenic diversity will therefore be determined by multiple factors. If all the selective forces are removed then the level of sequence diversity will remain the same. However, if there is a expansion of the viral population, then selection for replication rate will again favour the most fit viral variant with certain genetic and biological features, probably the same as those found in the early stage of infection. The reappearance of some early sequences in the later plasma samples of patient 82 may, to some extent, support this hypothesis (Chapter 4).

Apart from the viral quantitative elements which involved in the final breakdown of the immune system, there could also be some qualitative factors which hinder the normal functions of immune system. One of these factors could be the

changes in viral tropism. It has been known for some time that viral isolates obtained at later stages of infection are more T-cell-tropic rather than macrophage-tropic (Ross *et al.,* 1992; Schuitemaker *et al.,* 1992; Hwang *et al.,* 1991). In this aspect, viral evolution *in vivo* could therefore equally be described as the process of selection of T-cell-tropic virus variants with disease progression. This tropism shift to the T cells could be one of the reasons for the rapid reduction and dysfunctions of CD4+ cells in the blood circulation at the later stages of infection. Furthermore, the emergence of virus variants that are more virulent and cytopathic has been observed during the later stage of infection (Tersmette *et al.,* 1989*a, b*; Cheng-Mayer *et al.,* 1988). These results suggest that the development of disease symptoms in HIV-1 infected individuals could also be associated with the emergence of more T-cell tropic and pathogenic virus variants *in vivo*.

6.3 ORIGIN OF PLASMA VIRAEMIA Significant differences between the frequencies of sequence variants in plasma viral and lymphocyte-associated proviral population observed in this study indicated that at any one time point, the predominant plasma virus variants were antigenically distinct from those viruses encoded by HIV DNA sequences in PBMCs (see Chapter 4). The more rapid sequence turnover of viral RNA sequences than of proviral DNA sequences suggests that the new virus variants may appear and become dominant firstly in the plasma viral rather than in lymphocyte-associated proviral population. If this is the case, then it will be very reasonable to ask: where is the origin of plasma viraemia. There are several possible sources that could contribute to the plasma viraemia. First of all,

CD4+ T lymphocytes have been shown to be the major reservoir for HIV-1 in the peripheral blood compartment (Schnittman *et al.*, 1989). Quiescent as well as activated T lymphocytes can harbour HIV DNA, although proviral DNA may exist in different forms (see section 1.1.9). Quiescent T lymphocytes are able to produce HIV RNA as soon as they are activated by antigens or other stimulating factors (Zack *et al.*, 1990*a, b*; Stevenson *et al.*, 1990; Bukrinsky *et al.*, 1991), and therefore can become another major and inducible HIV reservoir in infected individuals. However, the low percentage of T lymphocytes (0.01 to 1%) that contain viral DNA and even lower percentage of these infected T lymphocytes (0.001-0.01%) actively expressing viral RNA is difficult to reconcile with the enormous amounts of plasma viraemia observed during the course of infection (Schnittman *et al.*, 1989; Simmonds *et al.*, 1990*a*).

HIV has been detected in several body tissues, such as those of lymph nodes, spleen, central nervous system, liver, bone marrow and lungs (Meltzer *et al.*, 1990*a, b, c*). In some of the patients studied, distinct virus variants have been found in different organs (Epstein *et al.*, 1991, Ball *et al.*, Regional Virus Laboratory, East Birmingham Hospital, Birmingham, personal communication). The uneven distribution of virus variants in the body may be due to the different biological features of virus variants present in these organs or due to other unknown factors. Examination of HIV-1 infected cells in these tissues demonstrated that the frequency of HIV-1 infected cells could be 10,000- to 100,000-fold higher than that in the blood compartment. The amount of HIV RNA per infected cell is at least 10-fold higher than that found in the blood CD4+ T lymphocytes (Meltzer *et al.*, 1990*a, b,*

c). The highly replicative nature of HIV-1 infected cells in these bodily tissues could therefore generate enormous amounts of virus variants which could then be shed into the plasma. If this is the case, the viruses present in different bodily tissues should be under more extensive study. These viruses as well as their host cells should be the target for the therapeutic strategies.

REFERENCES

Albert, J., B. Abrahamsson, K. Nagy, E. Aurelius, H. Gaines, G. Nystrom, and E. M. Fenyo. 1990. Rapid development of isolate-specific neutralizing antibodies after primary HIV-1 infection and consequent emergence of virus variants which resist neutralization by autologous sera. AIDS **4**:107-112.

Alizon, M., S. Wain Hobson, L. Montagnier, and P. Sonigo. 1986. Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. Cell **46**:63-74.

Allain, J. P., Y. Laurian, D. A. Paul, and D. Senn. 1986. Serological markers in early stages of human immunodeficiency virus infection in haemophiliacs. Lancet **2**:1233-1236.

Allan, J. S., M. Short, M. E. Taylor, S. Su, V. M. Hirsch, P. R. Johnson, G. M. Shaw, and B. H. Hahn. 1991. Species-specific diversity among simian immunodeficiency viruses from African green monkeys. J. Virol. **65**:2816-2828.

Al-Shawi, A., J. Burke, C. T. Jones, J. P. Simons and J. O. Bishop. 1988. A Mup promoter-thymidine kinase reporter gene shows relaxed tissue-specific expression and confer male sterility upon transgenic mice. Mol. Cell. Biol. **8**:4821-4828.

Answer (AIDS News Supplement, CDS Weekly Report). 1992. Human immunodeficiency virus type 1 (HIV-1) quarterly report to 30 June 1992.

Arya, S. K., C. Guo, S. F. Josephs, and F. Wong Staal. 1985. Trans-activator gene of human T-lymphotropic virus type III (HTLV-III). Science. **229**:69-73.

Baier, M., M. T. Dittmar, K. Cichutek, and R. Kurth. 1991. Development of vivo of genetic variability of simian immunodeficiency virus. Proc. Natl. Acad. Sci. U. S. A. **88**:8126-8130.

Baier, M., C. Garber, C. Muller, K. Cichutek, and R. Kurth. 1990. Complete nucleotide sequence of a simian immunodeficiency virus from African green monkeys: a novel type of intragroup divergence. Virology **176**:216-221.

Baier, M., A. Werner, K. Cichutek, C. Garber, C. Muller, G. Kraus, F. J. Ferdinand, S. Hartung, T. S. Papas, and R. Kurth. 1989. Molecularly cloned simian immunodeficiency virus SIVagm3 is highly divergent from other SIVagm isolates and is biologically active in vitro and in vivo. J. Virol. **63**:5119-5123.

Balfe, P., P. Simmonds, C. A. Ludlam, J. O. Bishop, and A. J. Leigh Brown. 1990. Concurrent evolution of human immunodeficiency virus type 1 in patients infected from the same source: rate of sequence change and low frequency of inactivating substitutions. J. Virol. **64**:6221-6233.

Barr, C. E., L. K. Miller, M. R. Lopez. 1992. Recovery of infectious HIV-1 from

whole saliva. J. Am. Dent. Assoc. **123**:37-45.

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

Bolognesi, D. P. 1990. Immunobiology of the human immunodeficiency virus envelope and its relationship to vaccine strategies. Mol. Biol. Med. **7**:1-15.

Brinchmann, J. E., J. Albert, and F. Vartdal. 1991. Few infected CD4+ T cells but a high proportion of replication-competent provirus copies in asymptomatic human immunodeficiency virus type 1 infection. J. Virol **65**:2019-2023.

Bukrinsky, M. I., T. L. Stanwick, M. P. Dempsey, and M. Stevenson. 1991. Quiescent T lymphocytes as an inducible virus reservoir in HIV-1 infection. Science. **254**:423-427.

Burns, D. P. and R. C. Desrosiers. 1991. Selection of genetic variants of simian immunodeficiency virus in persistently infected rhesus monkeys. J. Virol. **65**:1843-1854.

Byrne, B. C., J. J. Li, J. Sninsky, and B. J. Poiesz. 1988. Detection of HIV-1 RNA sequences by in vitro DNA amplification. Nucleic. Acids. Res. **16**:4165.

Cann, A. J., M. J. Churcher, M. Boyd, W. O'Brien, J. Q. Zhao, J. Zack, and I. S. Chen. 1992. The region of the envelope gene of human immunodeficiency virus type 1 responsible for determination of cell tropism. J. Virol. **66**:305-309.

Cease, K. B., H. Margalit, J. L. Cornette, S. D. Putney, W. G. Robey, C. Ouyang, H. Z. Streicher, P. J. Fischinger, R. C. Gallo, C. DeLisi, and et al. 1987. Helper T-cell antigenic site identification in the acquired immunodeficiency syndrome virus gp120 envelope protein and induction of immunity in mice to the native protein using a 16-residue synthetic peptide [published erratum appears in Proc Natl Acad Sci U S A 1988 Nov;85(21):8226]. Proc. Natl. Acad. Sci. U. S. A. **84**:4249-4253.

Center for Disease Control. 1986. Classification system for human t-lynphotropic virus type III/lymphadenopath-associated virus infection. MMWR. **35**: 334-339.

Center for Disease Control Statistic. 1992. AIDS. **6**:1229-1233.

Chakrabarti, L., M. Guyader, M. Alizon, M. D. Daniel, R. C. Desrosiers, P. Tiollais, and P. Sonigo. 1987. Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. Nature. **328**:543-547.

Chen, I. S., J. McLaughlin, J. C. Gasson, S. C. Clark, and D. W. Golde. 1983*a*. Molecular characterization of genome of a novel human T-cell leukaemia virus. Nature. **305**:502-505.

Chen, I. S., S. G. Quan, and D. W. Golde. 1983*b*. Human T-cell leukemia virus type II transforms normal human lymphocytes. Proc. Natl. Acad. Sci. U. S. A. **80**:7006-7009.

Cheng Mayer, C. and J. A. Levy. 1990*a*. Human immunodeficiency virus infection of the CNS: characterization of "neurotropic" strains. Curr. Top. Microbiol. Immunol. **160**:145-156.

Cheng Mayer, C., M. Quiroga, J. W. Tung, D. Dina, and J. A. Levy. 1990*b*. 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., J. T. Rutka, M. L. Rosenblum, T. McHugh, D. P. Stites, and J. A. Levy. 1987. Human immunodeficiency virus can productively infect cultured human glial cells. Proc. Natl. Acad. Sci. U. S. A. **84**:3526-3530.

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.

Cheng, S. M., L. Shaw-Guang, M. Ronchetti-Blume *et al.*,. 1991. Robust expression of the SIV envelope protein by a recombinant human adenovirus host-range mutant. In *Vaccines* 91: Modern Approaches to New Vaccines Including Prevention of AIDS. edited by Chanock R. M. *et al.*,. Cold Spring Harbor: Cold Spring Harbor Laboratory. 145-150.

Chesebro, B., K. Wehrly, J. Nishio, and S. Perryman. 1992. Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparsion with T-cell-tropic isolates: definition of critical amino acids involved in cell tropism. J. Virol **66**:6547-6554.

Chin, J. 1991. Global estimates of HIV infection and AIDS cases: 1991. AIDS. **5**:(suppl 2): s57-s61.

Chomczynski, P. and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162**:156-159.

Cichutek, K., S. Norley, R. Linde, W. Kreuz, M. Gahr, J. Lower, G. von Wangenheim, and R. Kurth. 1991. Lack of HIV-1 V3 region sequence diversity in two haemophiliac patients infected with a putative biologic clone of HIV-1. AIDS. **5**:1185-1187.

Clapham, P. R., A. Mcknight, and R. A. Weiss. 1992. Human immunodeficiency

virus type 2 infection and fusion of CD4-negative human cell lines: induction and enhancement by soluble CD4. J. Virol **66**:3531-3537.

Clark, S. J., M. S. Saag, W. D. Decker, S. Campbell Hill, J. L. Roberson, P. J. Veldkamp, J. C. Kappes, B. H. Hahn, and G. M. Shaw. 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., D. Guetard, F. Brun Vezinet, S. Chamaret, M. A. Rey, M. O. Santos Ferreira, A. G. Laurent, C. Dauguet, C. Katlama, C. Rouzioux, and et al. 1986*a*. Isolation of a new human retrovirus from West African patients with AIDS. Science. **233**:343-346.

Clavel, F., M. Guyader, D. Guetard, M. Salle, L. Montagnier, and M. Alizon. 1986*b*. Molecular cloning and polymorphism of the human immune deficiency virus type 2. Nature. **324**:691-695.

Clements, G. J., M. J. Price Jones, P. E. Stephens, C. Sutton, T. F. Schulz, P. R. Clapham, J. A. McKeating, M. O. McClure, S. Thomson, M. Marsh, and et al. 1991. The V3 loops of the HIV-1 and HIV-2 surface glycoproteins contain proteolytic cleavage sites: a possible function in viral fusion? AIDS Res. Hum. Retroviruses. **7**:3-16.

Clements, J. E., N. DAntonio, and O. Narayan. 1982. Genomic changes associated with antigenic variation of visna virus. II. Common nucleotide sequence changes detected in variants from independent isolations. J. Mol. Biol. **158**:415-434.

Clements, J. E., S. L. Gdovin, R. C. Montelaro, and O. Narayan. 1988. Antigenic Variation in Lentiviral Diseases. Ann. Rev. Immunol. **6**:139-159.

Clements, J. E., F. S. Pedersen, O. Narayan, and W. A. Haseltine. 1980. Genomic changes associated with antigenic variation of visna virus durig persistent infection. Proc. Natl. Acad. Sci. U. S. A. **77**:4454-4458.

Clerici, M., J. A. Berzofsky, G. M. Shearer, and C. O. Tacket. 1991. Exposure to human immunodeficiency virus (HIV) type I indicated by HIV-specific T helper cell responses before detection of infection by polymerase chain reaction and serum antibodies [corrected] [published erratum appears in J Infect Dis 1991 Oct; 164(4):832]. J. Infect. Dis. **164**:178-182.

Clerici, M., J. V. Giorgi, C. C. Chou, V. K. Gudeman, J. A. Zack, P. Gupta, H. N. Ho, P. G. Nishanian, J. A. Berzofsky, and G. M. Shearer. 1992. Cell-mediated immune response to human immunodeficiency virus (HIV) type 1 in seronegative homosexual men with recent sexual exposure to HIV-1. J. Infect. Dis. **165**:1012-1019.

Clumeck, N., H. Taelman, P. Hermans, P. Piot, M. Schoumacher, and S. De Wit.

1989. A cluster of HIV infection among heterosexual people without apparent risk factors. N. Engl. J. Med. **321**:1460-1462.

Coffin, J., A. Haase, J. A. Levy, L. Montagnier, S. Oroszlan, N. Teich, H. Temin, K. Toyoshima, H. Varmus, P. Vogt, and et al. 1986*a*. What to call the AIDS virus? Nature. **321**:10.

Coffin, J., A. Haase, J. A. Levy, L. Montagnier, S. Oroszlan, N. Teich, H. Temin, K. Toyoshima, H. Varmus, P. Vogt, and et al. 1986*b*. Human immunodeficiency viruses. Science. **232**:697.

Coffin, J. M. 1992. Genetic diversity and evolution of retroviruses. Curr. Top. Microbiol. Immunol. **176**:143-164.

Coombs, R. W., A. C. Collier, J. P. Allain, B. Nikora, M. Leuther, G. F. Gjerset, and L. Corey. 1989. Plasma viremia in human immunodeficiency virus infection [see comments]. N. Engl. J. Med. **321**:1626-1631.

Cooney, E. L., A. C. Collier, P. D. Greenberg *et al.*,. 1991. Safety of and immunological response to a recombinant vaccinia virus vaccine expressing HIV envelope glycoprotein. Lancet. **337**:567-572.

Cordonnier, A., L. Montagnier, and M. Emerman. 1989. Single amino-acid changes in HIV envelope affect viral tropism and receptor binding. Nature. **340**:571-574.

Crow, J. F., and M. Kimura. 1970. An introduction to population genetics theory (Harper and Row, New York).

Cullen, B. R. 1991. Regulation of HIV-1 gene expression. FASEB. J. **5**:2361-2368.

Daar, E. S., T. Moudgil, R. D. Meyer, and D. D. Ho. 1991. Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection. N. Engl. J. Med. **324**:961-964.

Dalgleish, A. G., P. C. Beverley, P. R. Clapham, D. H. Crawford, M. F. Greaves, and R. A. Weiss. 1984. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature. **312**:763-767.

Dalgleish, A. G., T. C. Chanh, R. C. Kennedy, P. Kanda, P. R. Clapham, and R. A. Weiss. 1988. Neutralization of diverse HIV-1 strains by monoclonal antibodies raised against a gp41 synthetic peptide. Virology. **165**:209-215.

De Jong, J. J., A. de Ronde, W. Keulen, M. Tersmette, and J. Goudsmit. 1992*a*. Minimal requirements for the human immunodeficiency virsu 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., J. Goudsmit, W. Keulen, B. Klaver, W. Krone, M. Tersmette, and A. de Ronde. 1992*b*. Human immunodeficiency virus type 1 clones chimeric for the envelope V3 domain differ in syncytium formation and replication capacity. J. Virol **66**:757-765.

Dedera, D., R. Gu, and L. Ratner. 1992. Conserved cysteine residues in the human immunodeficiency virus type 1 transmembrane envelope protein are essential for precursor envelope cleavage. J. Virol. **66**:1207-1209.

Desrosiers, R. C. 1988. Simian Immunodeficiency Viruses. Ann. Rev. Microbiol. **42**:607-625.

Desrosiers, R. C. 1990. A finger on the missing link. Nature **345**:288-289.

Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387-395.

Donahue, R. E., M. M. Johnson, L. I. Zon, S. C. Clark, and J. E. Groopman. 1987. Suppression of in vitro haematopoiesis following human immunodeficiency virus infection. Nature. **326**:200-203.

Earl, P. L., S. Koenig, B. Moss. Biological and immnological properties of human immunodeficiency virus type 1 envelope glycoprotein: analysis of proteins with truncations and deletions expressed by recombinant vaccinia viruses. J. Virol. **65**:31-41.

Edelman, A. S. and S. Zolla Pazner. 1989. AIDS: a syndrome of immune dysregulation, dysfunction, and deficiency. FASEB. J. **3**:22-30.

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 quasispecies. Virology **180**:583-590.

Fauci, A. S. 1988. The human immunodeficiency virus: infectivity and mechanisms of pathogenesis. Science. **239**:617-622.

Fauci, A. S. 1992. Immunopathogenic mechanisms of HIV infection. 7th Cent Gardes Meeting, France. 26-28th, October, 1992.

Feinberg, A. P. and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6-13.

Feinberg, M. B., R. F. Jarrett, A. Aldovini, R. C. Gallo, and F. Wong Staal.

1986. HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA. Cell. **46**:807-817.

Felsenstein, J. 1988. Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. **22**:521-565.

Feng, G., L. Yue, A. T. White, P. G. Pappas, J. Barchue, B. M. Greene, P. M. Sharp, G. M. Shaw, and B. H. Hahn. 1992. Human infection by genetically diverse SIVsm-related HIV-2 in West Africa. Nature **358**:495-499.

Feng, S. and E. C. Holland. 1988. HIV-1 tat trans-activation requires the loop sequence within tar. Nature **334**:165-167.

Fleischer, B. 1991*a*. T lymphocyte-stimulating microbial toxins as "superantigens". Med. Microbiol. Immunol. (Berl). **180**:53-58.

Fleischer, B. and H. W. Mittrucker. 1991*b*. Evidence for T cell receptor-HLA class II molecule interaction in the response to superantigenic bacterial toxins. Eur. J. Immunol. **21**:1331-1333.

Fouchier, R. A. M., M. Groenink, N. A. Kootstra, M. Tersmette, H. G. Huisman, F. Miedema, and H. Schuitemaker. 1992. Phenotype-associated sequence variation in the third variable domain of the Human Immunodeficiency virus type 1 gp120 molecule. J. Virol **66**:3138-3187.

Freed, E. O. and D. J. Myers. 1992. Identification and characterization of fusion and processing domains of the human immunodeficiency virus type 2 envelope glycoprotein. J. Virol **66**:5472-5478.

Freed, E. O., D. J. Myers, and R. Risser. 1991. Identification of the principal neutralizing determinant of human immunodeficiency virus type 1 as a fusion domain. J. Virol **65**:190-194.

Friedman Kien, A. E., L. J. Laubenstein, P. Rubinstein, E. Buimovici Klein, M. Marmor, R. Stahl, I. Spigland, K. S. Kim, and S. Zolla Pazner. 1982. Disseminated Kaposi's sarcoma in homosexual men. Ann. Intern. Med. **96**:693-700.

Fritsch, E. F. and H. M. Temin. 1977. Inhibition of viral DNA synthesis in stationary chicken embryo fibroblasts infected with avian retroviruses. J. Virol. **24**:461-469.

Gaines, H., J. Albert, M. von Sydow, A. Sonnerborg, F. Chiodi, A. Ehrnst, O. Strannegard, and B. Asjo. 1987. HIV antigenaemia and virus isolation from plasma during primary HIV infection. Lancet. **1**:1317-1318.

Gallaher, W. R. 1987. Detection of a fusion peptide sequence in the transmembrane

protein of human immunodeficiency virus. Cell. **50**:327-328.

Gallo, R. C., S. Z. Salahuddin, M. Popovic, G. M. Shearer, M. Kaplan, B. F. Haynes, T. J. Palker, R. Redfield, J. Oleske, B. Safai, 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.

Garson, J. A., P. W. Tuke, M. Makris, M. Briggs, S. J. Machin, F. E. Preston, and R. S. Tedder. 1990. Demonstration of viraemia patterns in haemophiliacs treated with hepatitis-C-virus-contaminated factor VIII concentrates. Lancet. **336**:1022-1025.

Gelderblom, H. R., E. H. Hausmann, M. Ozel, G. Pauli, and M. A. Koch. 1987. Fine structure of human immunodeficiency virus (HIV) and immunolocalization of structural proteins. Virology. **156**:171-176.

Gendelman, H. E. and M. S. Meltzer. 1989. Mononuclear phagocytes and the human immunodeficiency virus. Curr. Opin. Immunol. **2**:414-419.

Girard, M., and D. P.Bolognesi. 1991. Vaccines and immunology. AIDS. **5** (supple 2): S113-S114.

Goedert, J. J., R. J. Biggar, M. Melbye, D. L. Mann, S. Wilson, M. H. Gail, R. J. Grossman, R. A. DiGioia, W. C. Sanchez, S. H. Weiss, and et al. 1987. Effect of T4 count and cofactors on the incidence of AIDS in homosexual men infected with human immunodeficiency virus. JAMA. **257**:331-334.

Gonzalez Scarano, F., M. N. Waxham, A. M. Ross, and J. A. Hoxie. 1987. Sequence similarities between human immunodeficiency virus gp41 and paramyxovirus fusion proteins. AIDS. Res. Hum. Retroviruses. **3**:245-252.

Gottlieb, M. S., R. Schroff, H. M. Schanker, J. D. Weisman, P. T. Fan, R. A. Wolf, and A. Saxon. 1981. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. N. Engl. J. Med. **305**:1425-1431.

Goudsmit, J., N. K. Back, and P. L. Nara. 1991. Genomic diversity and antigenic variation of HIV-1: links between pathogenesis, epidemiology and vaccine development. FASEB. J. **5**:2427-2436.

Goudsmit, J., F. de Wolf, D. A. Paul, L. G. Epstein, J. M. Lange, W. J. Krone, H. Speelman, E. C. Wolters, J. Van der Noordaa, J. M. Oleske, and et al. 1986. Expression of human immunodeficiency virus antigen (HIV-Ag) in serum and cerebrospinal fluid during acute and chronic infection. Lancet. **2**:177-180.

Green, P. L. and I. S. Chen. 1990. Regulation of human T cell leukemia virus expression. FASEB. J. **4**:169-175.

Guy, B., M. P. Kieny, Y. Rivers, C. LePeuch, K. Dott, M. Girard, L. Montagnier, and J. P. Lecocq. 1987. HIV F/3' orf encodes a phosphorylated GTP-binding protein resembling an oncogene product. Nature. **330**:266-269.

Guyader, M., M. Emerman, P. Sonigo, F. Clavel, L. Montagnier, and M. Alizon. 1987. Genome organization and transactivation of the human immunodeficiency virus type 2. Nature. **326**:662-669.

Haase, A. T. 1986*a*. Pathogenesis of lentivirus infections. Nature. **322**:130-136.

Haase, A. T. 1986*b*. The pathogenesis of slow virus infections: molecular analyses. J. Infect. Dis. **153**:441-447.

Hahn, B. H., G. M. Shaw, S. K. Arya, M. Popovic, R. C. Gallo, and F. Wong Staal. 1984. Molecular cloning and characterization of the HTLV-III virus associated with AIDS. Nature. **312**:166-169.

Hahn, B. H., G. M. Shaw, M. E. Taylor, R. R. Redfield, P. D. Markham, S. Z. Salahuddin, F. Wong Staal, R. C. Gallo, E. S. Parks, and W. P. Parks. 1986. Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. Science **232**:1548-1553.

Haigwood, N. L., J. R. Shuster, G. K. Moore, H. Lee, P. V. Skiles, K. W. Higgins, P. J. Barr, C. George Nascimento, and K. S. Steimer. 1990. Importance of hypervariable regions of HIV-1 gp120 in the generation of virus neutralizing antibodies. AIDS. Res. Hum. Retroviruses. **6**:855-869.

Hammes, S. R., E. P. Dixon, M. H. Malim, B. R. Cullen, and W. C. Greene. 1989. Nef protein of human immunodeficiency virus type 1: evidence against its role as a transcriptional inhibitor. Proc. Natl. Acad. Sci. U. S. A. **86**:9549-9553.

Hansen, J. E., H. Clausen, C. Nielsen, L. S. Teglbjaerg, L. L. Hansen, C. M. Nielsen, E. Dabelsteen, L. Mathiesen, S. I. Hakomori, and J. O. Nielsen. 1990. Inhibition of human immunodeficiency virus (HIV) infection in vitro by anticarbohydrate monoclonal antibodies: peripheral glycosylation of HIV envelope glycoprotein gp120 may be a target for virus neutralization. J. Virol **64**:2833-2840.

Harper, M. E., M. H. Kaplan, L. M. Marselle, S. G. Pahwa, K. J. Chayt, M. G. Sarngadharan, F. Wong Staal, and R. C. Gallo. 1986. Concomitant infection with HTLV-I and HTLV-III in a patient with T8 lymphoproliferative disease. N. Engl. J. Med. **315**:1073-1078.

Hart, C., G. Schochetman, T. Spira, A. Lifson, J. Moore, J. Galphin, J. Sninsky, and C. Y. Ou. 1988. Direct detection of HIV RNA expression in seropositive subjects. Lancet **2**:596-599.

Hart, M. K., K. J. Weinhold, R. M. Scearce, E. M. Washburn, C. A. Clark, T. J. Palker, and B. F. Haynes. 1991. Priming of anti-human immunodeficiency virus (HIV) CD8+ cytotoxic T cells in vivo by carrier-free HIV synthetic peptides. Proc. Natl. Acad. Sci. U.S.A. **88**:9448-9452.

Hauber, J., A. Perkins, E. P. Heimer, and B. R. Cullen. 1987. Trans-activation of human immunodeficiency virus gene expression is mediated by nuclear events. Proc. Natl. Acad. Sci. U. S. A. **84**:6364-6368.

Helseth, E., U. Olshevsky, C. Furman, and J. Sodroski. 1991. Human immunodeficiency virus type 1 gp120 envelope glycoprotein important for association with the gp41 transmembrane glycoprotein. J. Virol. **65**:2119-2123.

Hirsch, V. M., G. Dapolito, C. McGann, R. A. Olmsted, R. H. Purcell, and P. R. Johnson. 1989*a*. Molecular cloning of SIV from sooty mangabey monkeys. J. Med. Primatol. **18**:279-285.

Hirsch, V. M., P. Edmondson, M. Murphey Corb, B. Arbeille, P. R. Johnson, and J. I. Mullins. 1989*b*. SIV adaptation to human cells. Nature. **19; 341**:573-574.

Hirsch, V. M., R. A. Olmsted, M. Murphey Corb, R. H. Purcell, and P. R. Johnson. 1989*c*. An African primate lentivirus (SIVsm) closely related to HIV-2. Nature. **339**:389-392.

Hirsch, V. M., P. M. Zack, and P. R. Johnson. 1990. Molecular characterization of SIV in tissues from experimentally infected macaques. J. Med. Primatol. **19**:287-294.

Ho, D. D., M. S. Fung, Y. Z. Cao, X. L. Li, C. Sun, T. W. Chang, and N. C. Sun. 1991*a*. Another discontinuous epitope on glycoprotein gp120 that is important in human immunodeficiency virus type 1 neutralization is identified by a monoclonal antibody. Proc. Natl. Acad. Sci. U. S. A. **88**:8949-8952.

Ho, D. D., J. A. McKeating, X. L. Li, T. Moudgil, E. S. Daar, N. C. Sun, and J. E. Robinson. 1991*b*. Conformational epitope on gp120 important in CD4 binding and human immunodeficiency virus type 1 neutralization identified by a human monoclonal antibody. J. Virol. **65**:489-493.

Ho, D. D., T. Moudgil, and M. Alam. 1989. Quantitation of human immunodeficiency virus type 1 in the blood of infected persons [see comments]. N. Engl. J. Med. **321**:1621-1625.

Ho, D. D., T. R. Rota, R. T. Schooley, J. C. Kaplan, J. D. Allan, J. E. Groopman, L. Resnick, D. Felsenstein, C. A. Andrews, and M. S. Hirsch. 1985. Isolation of HTLV-III from cerebrospinal fluid and neural tissues of patients with neurologic syndromes related to the acquired immunodeficiency syndrome. N. Engl. J. Med.

313:1493-1497.

Holmes, E. C., L. Q. Zhang, P. Simmonds, C. A. Ludlam, and A. J. Brown. 1992. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. Proc. Natl. Acad. Sci. U. S. A. **89**:4835-4839.

Homsy, J., M. Meyer, M. Tateno, S. Clarkson, and J. A. Levy. 1989. The Fc and not CD4 receptor mediates antibody enhancement of HIV infection in human cells. Science. **244**:1357-1360.

Hosmalin, A., M. Clerici, R. Houghten, C. D. Pendleton, C. Flexner, D. R. Lucey, B. Moss, R. N. Germain, G. M. Shearer, and J. A. Berzofsky. 1990. An epitope in human immunodeficiency virus 1 reverse transcriptase recognized by both mouse and human cytotoxic T lymphocytes. Proc. Natl. Acad. Sci. U. S. A. **87**:2344-2348.

Howell, R. M., J. E. Fitzgibbon, M. Noe, Z. J. Ren, D. J. Gocke, T. A. Schwartzer, and D. T. Dubin. 1991. In vivo sequence variation of the human immunodeficiency virus type 1 env gene: evidence for recombination among variants found in a single individual. AIDS. Res. Hum. Retroviruses. **7**:869-876.

Huet, T., R. Cheynier, A. Meyerhans, G. Roelants, and S. Wain Hobson. 1990. Genetic organization of a chimpanzee lentivirus related to HIV-1 [see comments]. Nature. **345**:356-359.

Hugin, A. W., M. S. Vacchio, and H. C. Morse. 1991. A virus-encoded "superantigen" in a retrovirus-induced immunodeficiency syndrome of mice. Science. **19; 252**:424-427.

Hwang, S. S., T. J. Boyle, H. K. Lyerly, and B. R. Cullen. 1991. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. Science **253**:71-74.

Imberti, L., A. Sottini, A. Bettinarde, M. Puoti, and D. Primi. 1991. Selective depletion in HIV infection of T cells that bear specific T cell receptor V beta sequences. Science **254**:860-862.

Jacks, T., K. Townsley, H. E. Varmus, and J. Majors. 1987. Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus gag-related polyproteins. Proc. Natl. Acad. Sci. U. S. A. **84**:4298-4302.

Jameson, B. A., P. E. Rao, L. I. Kong, B. H. Hahn, G. M. Shaw, L. E. Hood, and S. B. Kent. 1988. Location and chemical synthesis of a binding site for HIV-1 on the CD4 protein. Science. **240**:1335-1339.

Javaherian, K., A. J. Langlois, C. McDanal, K. L. Ross, L. I. Eckler, C. L. Jellis,

A. T. Profy, J. R. Rusche, D. P. Bolognesi, and S. D. Putney. 1989. Principal neutralizing domain of the human immunodeficiency virus type 1 envelope protein. Proc. Natl. Acad. Sci. U. S. A. **86**:6768-6772.

Johnson, P. R., A. Fomsgaard, J. Allan, M. Gravell, W. T. London, R. A. Olmsted, and V. M. Hirsch. 1990. Simian immunodeficiency viruses from African green monkeys display unusual genetic diversity. J. Virol. **64**:1086-1092.

Johnson, P. R., M. Gravell, J. Allan, S. Goldstein, R. A. Olmsted, G. Dapolito, C. McGann, W. T. London, R. H. Purcell, and V. M. Hirsch. 1989. Genetic diversity among simian immunodeficiency virus isolates from African green monkeys. J. Med. Primatol. **18**:271-277.

Johnson, P. R. and V. M. Hirsch. 1991. Pathogenesis of AIDS: the non-human primate model. AIDS **5 (suppl)**::s43-s48.

Johnson, P. R. and V. M. Hirsch. 1992. Genetic variation of simian immunodeficiency viruses in nonhuman primates. AIDS. Res. Hum. Retroviruses. **8**:367-372.

Joly, E., L. Mucke, M. B. A. Oldstone. 1991. Viral persistence in neurons explained by lack of major histocompatibility class I expression. Science. **253**:1283-1285.

Jones, K. A., J. T. Kadonga, P. A. Tijian. 1986. Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. Science. **232**:755-759.

Jouault, T., F. Chapuis, R. Olivier, C. Parravicini, E. Bahraoui, and J. C. Gluckman. 1989. HIV infection of monocytic cells: role of antibody-mediated virus binding to Fc-gamma receptors. AIDS. **3**:125-133.

Jurriaans, S., J. T. Dekker, and A. D. Ronde. 1992. HIV-1 viral DNA load in peripheral blood monocuclear cells from seroconverters and long-term infected individuals. AIDS **6**:635-641.

Kellam, P., C. A. Boucher, and B. A. Larder. 1992. Fifth mutation in human immunodeficiency virus type 1 reverse transcriptase contributes to the development of high-level resistance to zidovudine. Proc. Natl. Acad. Sci. U. S. A. **89**:1934-1938.

Kestler, H. W., D. J. Ringler, K. Mori, D. L. Panicali, P. K. Sehgal, M. D. Daniel, and R. C. Desrosiers. 1991. Importance of the nef gene for maintenance of high virus loads and for development of AIDS. Cell. **65**:651-662.

Klatzmann, D., E. Champagne, S. Chamaret, J. Gruest, D. Guetard, T. Hercend, J. C. Gluckman, and L. Montagnier. 1984. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. Nature. **312**:767-768.

Koenig, S., H. E. Gendelman, J. M. Orenstein, M. C. Dal Canto, G. H. Pezeshkpour, M. Yungbluth, F. Janotta, A. Aksamit, M. A. Martin, and A. S. Fauci. 1986. Detection of AIDS virus in macrophages in brain tissue from AIDS patients with encephalopathy. Science. **233**:1089-1093.

Kraus, G., A. Werner, M. Baier, D. Binniger, F. J. Ferdinand, S. Norley, and R. Kurth. 1989. Isolation of human immunodeficiency virus-related simian immunodeficiency viruses from African green monkeys. Proc. Natl. Acad. Sci. U. S. A. **86**:2892-2896.

Kusumi, K., B. Conway, S. Cunningham, A. Berson, C. Evans, A. K. Iversen, D. Colvin, M. V. Gallo, S. Coutre, E. G. Shpaer, and et al. 1992. Human immunodeficiency virus type 1 envelope gene structure and diversity in vivo and after cocultivation in vitro. J. Virol. **66**:875-885.

Laga, M., H. Taelman, P. Van der Stuyft, L. Bonneux, G. Vercauteren, and P. Piot. 1989. Advanced immunodeficiency as a risk factor for heterosexual transmission of HIV. AIDS **3**:361-366.

Lange, J. M., D. A. Paul, F. de Wolf, R. A. Coutinho, and J. Goudsmit. 1987. Viral gene expression, antibody production and immune complex formation in human immunodeficiency virus infection. AIDS **1**:15-20.

Larder, B. A., G. Darby, and D. D. Richman. 1989. HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy. Science. **243**:1731-1734.

Larder, B. A. and S. D. Kemp. 1989. Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). Science. **246**:1155-1158.

Larder, B. A., S. D. Kemp, and D. J. Purifoy. 1989. Infectious potential of human immunodeficiency virus type 1 reverse transcriptase mutants with altered inhibitor sensitivity. Proc. Natl. Acad. Sci. U. S. A. **86**:4803-4807.

LaRosa, G. J., J. P. Davide, K. Weinhold, J. A. Waterbury, A. T. Profy, J. A. Lewis, A. J. Langlois, G. R. Dreesman, R. N. Boswell, P. Shadduck, and et al. 1990. Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant [published erratum appears in Science 1991 Feb 15;251(4995):811]. Science. **249**:932-935.

Lasky, L. A., G. Nakamura, D. H. Smith, C. Fennie, C. Shimasaki, E. Patzer, P. Berman, T. Gregory, and D. J. Capon. 1987. Delineation of a region of the human immunodeficiency virus type 1 gp120 glycoprotein critical for interaction with the CD4 receptor. Cell. **50**:975-985.

Leigh Brown, A. and P. Monaghan. 1988. Evolution of the structural proteins of

human immunodeficiency virus: selective constraints on nucleotide substitution. AIDS Res. Hum. Retroviruses **4**:399-407.

Leigh Brown, A. J. 1991. Sequence variability in the human immunodeficiency viruses: pattern and process in viral evolution, M. W. Adler, J. W. M. Gold, and J. A. Levy (eds.), AIDS '91 - A year in review. Current Science, London.

Levy, J. A., J. Shimabukuro, T. McHugh, C. Casavant, D. Stites, and L. Oshiro. 1985. AIDS-associated retroviruses (ARV) can productively infect other cells besides human T helper cells. Virology. **147**:441-448.

Li, W. H., M. Tanimura, and P. M. Sharp. 1988. Rates and dates of divergence between AIDS virus nucleotide sequences. Mol. Biol. Evol. **5**:313-330.

Lifson, J. D. and E. G. Engleman. 1989. Role of CD4 in normal immunity and HIV infection. Immunol. Rev. **109**:93-117.

Lifson, J. D., M. B. Feinberg, G. R. Reyes, L. Rabin, B. Banapour, S. Chakrabarti, B. Moss, F. Wong Staal, K. S. Steimer, and E. G. Engleman. 1986*a*. Induction of CD4-dependent cell fusion by the HTLV-III/LAV envelope glycoprotein. Nature. **323**:725-728.

Lifson, J. D., G. R. Reyes, M. S. McGrath, B. S. Stein, and E. G. Engleman. 1986*b*. AIDS retrovirus induced cytopathology: giant cell formation and involvement of CD4 antigen. Science. **232**:1123-1127.

Looney, D. J., A. G. Fisher, S. D. Putney, J. R. Rusche, R. R. Redfield, D. S. Burke, R. C. Gallo, and F. Wong Staal. 1988. Type-restricted neutralization of molecular clones of human immunodeficiency virus. Science. **241**:357-359.

Ludlam, C. A., J. Tucker, C. M. Steel, R. S. Tedder, R. Cheingsong Popov, R. A. Weiss, D. B. McClelland, I. Philp, and R. J. Prescott. 1985. Human T-lymphotropic virus type III (HTLV-III) infection in seronegative haemophiliacs after transfusion of factor VIII. Lancet. **2**:233-236.

Lyerly, H. K., T. J. Matthews, A. J. Langlois, D. P. Bolognesi, and K. J. Weinhold. 1987*a*. Human T-cell lymphotropic virus IIIB glycoprotein (gp120) bound to CD4 determinants on normal lymphocytes and expressed by infected cells serves as target for immune attack. Proc. Natl. Acad. Sci. U. S. A. **84**:4601-4605.

Lyerly, H. K., D. L. Reed, T. J. Matthews, A. J. Langlois, P. A. Ahearne, S. R. Petteway, and K. J. Weinhold. 1987*b*. Anti-GP 120 antibodies from HIV seropositive individuals mediate broadly reactive anti-HIV ADCC. AIDS. Res. Hum. Retroviruses. **3**:409-422.

Maddon, P. J., A. G. Dalgleish, J. S. McDougal, P. R. Clapham, R. A. Weiss, and

R. Axel. 1986. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. Cell. **47**:333-348.

Malim, M. H., S. Bohnlein, J. Hauber, and B. R. Cullen. 1989*a*. Functional dissection of the HIV-1 Rev trans-activator--derivation of a trans-dominant repressor of Rev function. Cell. **58**:205-214.

Malim, M. H., L. Hauber, SY. Le, J. V. Maizel, and B. R. Cullen. 1989*b*. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature **338**:254-257.

Maniatis, T., E. F. Fritsch, J. Sambrook. 1982. Molecular Cloning (A laboratory manual). Cold Spring Harbor Laboratory.

Masur, H., M. A. Michelis, J. B. Greene, I. Onorato, R. A. Stouwe, R. S. Holzman, G. Wormser, L. Brettman, M. Lange, H. W. Murray, and S. Cunningham Rundles. 1981. An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. N. Engl. J. Med. **305**:1431-1438.

McCune, J. M., R. Namikawa, C. C. Shih, L. Rabin, and H. Kaneshima. 1990. Suppression of HIV infection in AZT-treated SCID-hu mice. Science. **247**:564-566.

McElrath, M. J., J. E. Pruett, and Z. A. Cohn. 1989. Mononuclear phagocytes of blood and bone marrow: comparative roles as viral reservoirs in human immunodeficiency virus type 1 infections. Proc. Natl. Acad. Sci. U. S. A. **86**:675-679.

McKeating, J. A., J. Gow, J. Goudsmit, L. H. Pearl, C. Mulder, and R. A. Weiss. 1989. Characterization of HIV-1 neutralization escape mutants. AIDS **3**:777-784.

McKeating, J. A., P. D. Griffiths, and R. A. Weiss. 1990. HIV susceptibility conferred to human fibroblasts by cytomegalovirus-induced Fc receptor. Nature. **343**:659-661.

McRae, B., J. A. Lange, M. S. Ascher, F. de Wolf, H. W. Sheppard, J. Goudsmit, and J. P. Allain. 1991. Immune response to HIV p24 core protein during the early phases of human immunodeficiency virus infection. AIDS Res. Hum. Retroviruses. **7**:637-643.

Meltzer, M. S., M. Nakamura, B. D. Hansen, J. A. Turpin, D. C. Kalter, and H. E. Gendelman. 1990*a*. Macrophages as susceptible targets for HIV infection, persistent viral reservoirs in tissue, and key immunoregulatory cells that control levels of virus replication and extent of disease. AIDS. Res. Hum. Retroviruses. **6**:967-971.

Meltzer, M. S., D. R. Skillman, P. J. Gomatos, D. C. Kalter, and H. E. Gendelman. 1990*b*. Role of mononuclear phagocytes in the pathogenesis of human

immunodeficiency virus infection. Annu. Rev. Immunol. **8**:169-194.

Meltzer, M. S., D. R. Skillman, D. L. Hoover, B. D. Hanson, J. A. Turpin, D. C. Kalter, and H. E. Gendelman. 1990*c*. Macrophages and the human immunodeficiency virus. Immunol. Today. **11**:217-223.

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

Minor, P. D. 1989. Strategies for the development of vaccine agaist HIV. J. antimicorbial chemotherapy. **23**, suppl. A, 55-62.

Mizuma, H., S. Litwin, S. Zolla-Pazner. 1988. B cell activation in HIV infection: Relationship of spontaneous immunoglobulin secretion to various immunologic parameters. Clin. Exp. Immunol. **71**:410-416.

Modrow, S., B. H. Hahn, G. M. Shaw, R. C. Gallo, F. Wong Staal, and H. Wolf. 1987. Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: prediction of antigenic epitopes in conserved and variable regions. J. Virol. **61**:570-578.

Mohri, H., M. K. Singh, W. T. W. Ching, and D. D. Ho. 1993. Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in the blood of treated and untreated patients. Porc. Natl. Acad. Sci. U. S. A. **90**:25-29.

Montefiori, D. C., W. E. Robinson,Jr., and W. M. Mitchell. 1988. Role of protein N-glycosylation in pathogenesis of human immunodeficiency virus type 1. Proc. Natl. Acad. Sci. U. S. A. **85**:9248-9252.

Montefiori, D. C., I. Y. Zhou, B. Barnes, D. Lake, E. M. Hersh, Y. Masuho, and L. B. Lefkowitz. 1991. Homotypic antibody responses to fresh clinical isolates of human immunodeficiency virus. Virology. **182**:635-643.

Moore, J. P., J. A. McKeating, R. A. Weiss, and Q. J. Sattentau. 1990. Dissociation of gp120 from HIV-1 virions induced by soluble CD4. Science. **250**:1139-1142.

Myers, G., B. Korber, J. Berzofsky, T. F. Smith, and G. N. Pavlakis. 1991. Human retroviruses and AIDS 1991. Los Alamos National Laboratory, Los Alamos, New Mexico.

Nabel, G., S. A. Rice, D. M. Knipe, D. Baltimore. 1988. Alternative mechanism for activation of human immunodeficiency virus enhancer in T cells. Science. **239**: 1299-1302.

Navia, B. A., E. S. Cho, C. K. Petito, and R. W. Price. 1986*a*. The AIDS dementia complex: II. Neuropathology. Ann. Neurol. **19**:525-535.

Navia, B. A., B. D. Jordan, and R. W. Price. 1986*b*. The AIDS dementia complex: I. Clinical features. Ann. Neurol. **19**:517-524.

Nelson, J. A., P. Ghazal, and C. A. Wiley. 1990. Role of opportunistic viral infections in AIDS. AIDS. **4**:1-10.

Nicholson, J. K., G. D. Cross, C. S. Callaway, and J. S. McDougal. 1986. In vitro infection of human monocytes with human T lymphotropic virus type III/lymphadenopathy-associated virus (HTLV-III/LAV). J. Immunol. **137**:323-329.

Niederman, T. M., B. J. Thielan, and L. Ratner. 1989. Human immunodeficiency virus type 1 negative factor is a transcriptional silencer. Proc. Natl. Acad. Sci. U. S. A. **86**:1128-1132.

Nishanian, P., K. R. Huskins, S. Stehn, R. Detels, and J. L. Fahey. 1990. A simple method for improved assay demonstrates that HIV p24 antigen is present as immune complexes in most sera from HIV-infected individuals. J. Infect. Dis. **162**:21-28.

Nowak, M. A., R. M. Anderson, A. R. McLean, T. F. Wolfs, J. Goudsmit, and R. M. May. 1991*a*. Antigenic diversity thresholds and the development of AIDS [see comments]. Science. **254**:963-969.

Nowak, M. A. and R. M. May. 1991*b*. Mathematical biology of HIV infections: antigenic variation and diversity threshold. Math. Biosci. **106**:1-21.

Nowak, M. A., R. M. May, and R. M. Anderson. 1990. The evolutionary dynamics of HIV-1 quasispecies and the development of immunodeficiency disease. AIDS. **4**:1095-1103.

O'Brien, W. A., Y. Koyanagi, A. Namazie, J. Q. Zhao, A. Diagne, K. Idler, J. A. Zack, and I. S. Chen. 1990. HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120 outside the CD4-binding domain. Nature. **348**:69-73.

Oka, S., K. Urayama, Y. Hirabayashi, K. Ohnishi, H. Goto, K. Mitamura, S. Kimura, and K. Shimada. 1990. Quantitative analysis of human immunodeficiency virus type-1 DNA in asymptomatic carriers using the polymerase chain reaction. Biochem. Biophys. Res. Commun. **167**:1-8.

Ou, C. Y., S. Kwok, S. W. Mitchell, D. H. Mack, J. J. Sninsky, J. W. Krebs, P. Feorino, D. Warfield, and G. Schochetman. 1988. DNA amplification for direct detection of HIV-1 in DNA of peripheral blood mononuclear cells [published erratum appears in Science 1988 Apr 8;240(4849):240]. Science. **239**:295-297.

Pahwa, S., R. Pahwa, C. Saxinger, R. C. Gallo, and R. A. Good. 1985. Influence of the human T-lymphotropic virus/lymphadenopathy-associated virus on functions of human lymphocytes: evidence for immunosuppressive effects and polyclonal B-cell activation by banded viral preparations. Proc. Natl. Acad. Sci. U. S. A. **82**:8198-8202.

Palker, T. J., M. E. Clark, A. J. Langlois, T. J. Matthews, K. J. Weinhold, R. R. Randall, D. P. Bolognesi, and B. F. Haynes. 1988. Type-specific neutralization of the human immunodeficiency virus with antibodies to env-encoded synthetic peptides. Proc. Natl. Acad. Sci. U. S. A. **85**:1932-1936.

Pang, S., Y. Schlesinger, E. S. Daar, T. Moudgil, D. D. Ho and I. S. Y. Chen. 1992. Rapid generation of sequence variation during primary HIV-1 infection. AIDS. **6**:453-460.

Papsidero, L. D., M. Sheu, and F. W. Ruscetti. 1989. Human immunodeficiency virus type 1-neutralizing monoclonal antibodies which react with p17 core protein: characterization and epitope mapping. J. Virol **63**:267-272.

Phillips, A. N. 1992. Studies of prognostic markers in HIV infection: implications for pathogenesis. AIDS. **6**:1391-1394.

Phillips, R. E., S. Rowland-Jones, D. F. Nixon, F. M. Gotch, J. P. Edwards, A. O. Ogunlesi, J. G. Elvin, J. A. Rothbard, C. R. M. Banghman, C. R. Rizza, and A. J. McMichael. 1991. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. Nature **354**:453-459.

Poli, G., B. Bottazzi, R. Acero, L. Bersani, V. Rossi, M. Introna, A. Lazzarin, M. Galli, and A. Mantovani. 1985. Monocyte function in intravenous drug abusers with lymphadenopathy syndrome and in patients with acquired immunodeficiency syndrome: selective impairment of chemotaxis. Clin. Exp. Immunol. **62**:136-142.

Popovic, M., M. G. Sarngadharan, E. Read, and R. C. Gallo. 1984. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science. **224**:497-500.

Preston, B. D., B. J. Poiesz, and L. A. Loeb. 1988. Fidelity of HIV-1 reverse transcriptase. Science **242**:1168-1171.

Prevec, L., B. S. Christie, K. E. Laurie *et al.*,. Immune response to HIV-1 Gag antigens induced by recombiant adenovirus vectors in mice and rhesus macaque monkeys. 1991. J. Acquir. Immune. Defic. Syndr. **4**:5.68-576

Price, R. W., B. A. Navia, and E. S. Cho. 1986. AIDS encephalopathy. Neurol. Clin. **4**:285-301.

Profy, A. T., P. A. Salinas, L. I. Eckler, N. M. Dunlop, P. L. Nara, and S. D.

Putney. 1990. Epitopes recognized by the neutralizing antibodies of an HIV-1-infected individual. J. Immunol. **144**:4641-4647.

Putney, S. D. and J. A. McKeating. 1990*a*. Antigenic variation in HIV. AIDS **4**:(suppl): s129-s136.

Putney, S. D., J. Rusche, K. Javaherian, T. Matthews, and D. Bolognesi. 1990*b*. Structural and functional features of the HIV envelope glycoprotein and considerations for vaccine development. Biotechnology. **14**:81-110.

Ragni, M. V., L. A. Kingsley, P. Nimorwicz, P. Gupta, and C. R. Rinaldo. 1989. HIV heterosexual transmission in hemophilia couples: lack of relation to T4 number, clinical diagnosis, or duration of HIV exposure. J. Acquir. Immune. Defic. Syndr. **2**:557-563.

Ratner, L., W. Haseltine, R. Patarca, K. J. Livak, B. Starcich, S. F. Josephs, E. R. Doran, J. A. Rafalski, E. A. Whitehorn, K. Baumeister, and et al. 1985. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature. **313**:277-284.

Redfield, R. R., D. C. Wright, E. C. Tramont. 1986. The Aalter Reed staging classification for HTLV-III/LAV infection. N. Engl. J. Med. **314**:131-132.

Richman, D. D. 1992. HIV drug resistance. AIDS Res. Hum. Retroviruses. **8**:1065-1071.

Roberts, J. D., K. Bebenek, and T. A. Kunkel. 1988. The accuracy of reverse transcriptase from HIV-1. Science. **242**:1171-1173.

Rosen, C. A., J. G. Sodroski, and W. A. Haseltine. 1985. The location of cis-acting regulatory sequences in the human T cell lymphotropic virus type III (HTLV-III/LAV) long terminal repeat. Cell. **41**:813-823.

Ross, M. T. L., Lange J.M.A., R. E. Y. Goede, R. A. Coutinho, P. T. A. Schellekens, F. Miedema, and M Tersmette. 1992. Viral Phenotype and Immune Response in Primary Human Immunodeficiency Virus Type 1 Infection. J. Infect. Dis. **165**:427-432.

Rusche, J. R., K. Javaherian, C. McDanal, J. Petro, D. L. Lynn, R. Grimaila, A. Langlois, R. C. Gallo, L. O. Arthur, P. J. Fischinger, and et al. 1988. Antibodies that inhibit fusion of human immunodeficiency virus-infected cells bind a 24-amino acid sequence of the viral envelope, gp120 [published errata appear in Proc Natl Acad Sci U S A 1988 Nov;85(22):8697 and Proc Natl Acad Sci U S A 1989 Mar; 86(5):1667]. Proc. Natl. Acad. Sci. U. S. A. **85**:3198-3202.

Ryu, S., P. D. Kwong, A. Truneh, T. G. Porter, J. Arthos, M. Rosenberg, X. Dai, N. Xuong, R, Axel, R. W. Sweet and W. A. Hendrickson. Crystal structure of an

HIV-binding recombinant fragment of human CD4. 1990. Nature. **348**:419-426.

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science. **239**:487-491.

Sattentau, Q. J. and J. P. Moore. 1991. Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. J. Exp. Med. **174**:407-415.

Schnittman, S. M., J. J. Greenhouse, H. C. Lane, P. F. Pierce, and A. S. Fauci. 1991. Frequent detection of HIV-1-specific mRNAs in infected individuals suggests ongoing active viral expression in all stages of disease. AIDS Res. Hum. Retroviruses **7**:361-367.

Schnittman, S. M., H. C. Lane, J. Greenhouse, J. S. Justement, M. Baseler, and A. S. Fauci. 1990. Preferential infection of CD4+ memory T cells by human immunodeficiency virus type 1: evidence for a role in the selective T-cell functional defects observed in infected individuals. Proc. Natl. Acad. Sci. U. S. A. **87**:6058-6062.

Schnittman, S. M., M. C. Psallidopoulos, H. C. Lane, L. Thompson, M. Baseler, F. Massari, C. H. Fox, N. P. Salzman, and A. S. Fauci. 1989. The reservoir for HIV-1 in human peripheral blood is a T cell that maintains expression of CD4 [published erratum appears in Science 1989 Aug 18;245(4919):preceding 694]. Science. **245**:305-308.

Schuitemaker, H., M. Koot, N. Kootstra, M. W. Dercksen, R. T. D. Goede, R. Steenwijk, J. M. A. Lange, J. K. Schatttenkerk, F. Miedema, and M. Tersmette. 1992. Biologial 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.

Schuitemaker, H., N. A. Kootstra, R. E. de Goede, F. de Wolf, F. Miedema, and M. Tersmette. 1991. Monocytotropic human immunodeficiency virus type 1 (HIV-1) variants detectable in all stages of HIV-1 infection lack T-cell line tropism and syncytium-inducing ability in primary T-cell culture. J. Virol. **65**:356-363.

Sharma, D. P., M. C. Zink, M. Anderson, R. Adams, J. E. Clements, S. V. Joag and O. Narayan. 1992. Derivation of neurotropic simmain immunodeficiency virus from exclusively lymphocytetropic parental virus: pathogenesis of infection in Macaques. **66**:3550-3556.

Sharp, P. M. and W. Li. 1988. Understanding the origin of AIDS viruses. Nature **336**:315.

Shaw, G. M., B. H. Hahn, S. K. Arya, J. E. Groopman, R. C. Gallo, and F. Wong Staal. 1984. Molecular characterization of human T-cell leukemia (lymphotropic) virus type III in the acquired immune deficiency syndrome. Science. **226**:1165-1171.

Shen, L., Z. W. Chen, M. D. Miller *et al.*,. 1991. Recombinant virus vaccine0induced SIV-specific CD8+ cytotoxic T-lymphocytes. Science. **252**:440-443.

Simmonds, P., P. Balfe, C. A. Ludlam, J. O. Bishop, and A. J. Brown. 1990*a*. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. J. Virol. **64**:5840-5850.

Simmonds, P., F. A. L. Lainson, R. Cuthbert, C. M. Steel, J. F. Peutherer, and C. A. Ludlam. 1988. HIV antigen and antibody detection: variable responses to infection in the Edinburgh haemophiliac cohort. British Medical Journal **296**:593-598.

Simmonds, P., L. Q. Zhang, F. McOmish, P. Balfe, C. A. Ludlam, and A. J. Leigh Brown. 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.

Simmonds, P., L. Q. Zhang, H. G. Watson, S. Rebus, E. D. Ferguson, P. Balfe, G. H. Leadbetter, P. L. Yap, J. F. Peutherer, C. A. Ludlum. 1990. Hepatitis C quantification and sequencing in blood products, haemophiliacs, and drug users. Lancet. **336**:1469-1472.

Smith, T. F., A. Srinivasan, G. Schochetman, M. Marcus, and G. Myers. 1988. The phylogenetic history of immunodeficiency viruses. Nature **333**:573-575.

Sodroski, J., W. C. Goh, C. Rosen, K. Campbell, and W. A. Haseltine. 1986*a*. Role of the HTLV-III/LAV envelope in syncytium formation and cytopathicity. Nature. **322**:470-474.

Sodroski, J., W. C. Goh, C. Rosen, A. Dayton, E. Terwilliger, and W. Haseltine. 1986*b*. A second post-transcriptional trans-activator gene required for HTLV-III replication. Nature. **321**:412-417.

Sodroski, J., R. Patarca, C. Rosen, F. Wong Staal, and W. Haseltine. 1985*a*. Location of the trans-activating region on the genome of human T-cell lymphotropic virus type III. Science. **229**:74-77.

Sodroski, J. G., C. A. Rosen, and W. A. Haseltine. 1984. Trans-acting transcriptional activation of the long terminal repeat of human T lymphotropic viruses in infected cells. Science. **225**:381-385.

Sodroski, J., C. Rosen, F. Wong Staal, S. Z. Salahuddin, M. Popovic, S. Arya, R. C. Gallo, and W. A. Haseltine. 1985*b*. Trans-acting transcriptional regulation of human T-cell leukemia virus type III long terminal repeat. Science. **227**:171-173.

St Clair, M. H., J. L. Martin, G. Tudor Williams, M. C. Bach, C. L. Vavro, D. M. King, P. Kellam, S. D. Kemp, and B. A. Larder. 1991. Resistance to ddI and sensitivity to AZT induced by a mutation in HIV-1 reverse transcriptase. Science. **253**:1557-1559.

Stahl, R. E., A. Friedman Kien, R. Dubin, M. Marmor, and S. Zolla Pazner. 1982. Immunologic abnormalities in homosexual men. Relationship to Kaposi's sarcoma. Am. J. Med. **73**:171-178.

Starcich, B. R, B. H. Hahn, G. M. Shaw, P. D. McNeely, S. Modrow, H. Wolf, E. S. Parks, W. P. Parks, S. F Josephs, R. C. Gallo, and F. Wong Staal. 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.

Steel, C. M., C. A. Ludlam, D. Beatson, J. F. Peutherer, R. J. G. Cuthbert, P. Simmonds, M. Morrison, and M. Jones. 1988. HLA haplotype A1 B8 DR3 as a risk factor for HIV-related disease. Lancet. **May 28**:1185-1188.

Steimer, K. S., C. J. Scandella, P. V. Skiles, and N. L. Haigwood. 1991. Neutralization of divergent HIV-1 isolates by conformation-dependent human antibodies to Gp120. Science. **254**:105-108.

Stein, B. S. and E. G. Engleman. 1990. Intracellular processing of the gp160 HIV-1 envelope precursor. Endoproteolytic cleavage occurs in a cis or medial compartment of the Golgi complex. J. Biol. Chem. **265**:2640-2649.

Stein, B. S., S. D. Gowda, J. D. Lifson, R. C. Penhallow, K. G. Bensch, and E. G. Engleman. 1987. pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane. Cell. **49**:659-668.

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.

Stevenson, M., T. L. Stanwick, M. P. Dempsey, and C. A. Lamonica. 1990. HIV-1 replication is controlled at the level of T cell activation and proviral integration. EMBO. J. **9**:1551-1560.

Takahashi, H., J. Cohen, A. Hosmalin, K. B. Cease, R. Houghten, J. L. Cornette, C. DeLisi, B. Moss, R. N. Germain, and J. A. Berzofsky. 1988. An immunodominant epitope of the human immunodeficiency virus envelope glycoprotein gp160 recognized by class I major histocompatibility complex

molecule-restricted murine cytotoxic T lymphocytes. Proc. Natl. Acad. Sci. U. S. A. **85**:3105-3109.

Takeuchi, Y., M. Akutsu, K. Murayama, N. Shimizu, and H. Hoshino. 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.

Tersmette, M., R. A. Gruters, F. de Wolf, R. E. de Goede, J. M. Lange, P. T. Schellekens, J. Goudsmit, H. G. Huisman, and F. Miedema. 1989*a*. Evidence for a role of virulent human immunodeficiency virus (HIV) variants in the pathogenesis of acquired immunodeficiency syndrome: studies on sequential HIV isolates. J. Virol. **63**:2118-2125.

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*b*. Association between biological properties of human immunodeficiency virus variants and risk for AIDS and AIDS mortality. Lancet **1**:983-985.

Tindall, B., L. Evans, P. Cunningham, P. McQueen, L. Hurren, E. Vasak, J. Mooney, and D. A Cooper. 1992. Indentification of HIV-1 in semen following primary HIV-1 infection. AIDS **6**:949-952.

Tschachler, E., H. Buchow, R. C. Gallo, and M. S. Reitz. 1990. Functional contribution of cysteine residues to the human immunodeficiency virus type 1 envelope. J. Virol. **64**:2250-2259.

Ujhelyi, E., B. Buki, V. Salavecz, D. Banhegyi, A. Horvath, G. Fust, and S. R. Hollan. 1987. A simple method for detecting HIV antibodies hidden in circulating immune complexes. AIDS **1**:161-165.

Varmus, H. 1988*a*. Retroviruses. Science **240**:1427-1435.

Varmus, H. 1988*b*. Regulation of HIV and HTLV gene expression. Gene & Development **2**:1055-1062.

Varmus, H. E., T. Padgett, S. Heasley, G. Simon, and J. M. Bishop. 1977. Cellular functions are required for the synthesis and integration of avian sarcoma virus-specific DNA. Cell. **11**:307-319.

Veronese, F. D., T. D. Copeland, S. Oroszlan, R. C. Gallo, and M. G. Sarngadharan. 1988. Biochemical and immunological analysis of human immunodeficiency virus gag gene products p17 and p24. J. Virol **62**:795-801.

Veronese, F. D., A. L. DeVico, T. D. Copeland, S. Oroszlan, R. C. Gallo, and M.

G. Sarngadharan. 1985. Characterization of gp41 as the transmembrane protein coded by the HTLV-III/LAV envelope gene. Science. **229**:1402-1405.

Wain-Hobson, S., P. Sonigo, O. Danos, S. Cole, and M. Alizon. 1985. Nucleotide sequence of the AIDS virus, LAV. Cell. **40**:9-17.

Wain-Hobson, S., J. P. Vartanian, M. Henry, N. Chenciner, R. Cheynier, S. Delassus, L. P. Martins, M. Sala, M. T. Nugeyre, D. Guetard, et al. 1991. LAV revisited: Origins of the early HIV-1 isolates from Institut Pasteur. Science. **252**:961-965.

Walker, B. D., C. Flexner, K. Birch Limberger, L. Fisher, T. J. Paradis, A. Aldovini, R. Young, B. Moss, and R. T. Schooley. 1989. Long-term culture and fine specificity of human cytotoxic T-lymphocyte clones reactive with human immunodeficiency virus type. Proc. Natl. Acad. Sci. U. S. A. **86**:9514-9518.

Wang, J., Y. Yan, T. P. J. Garrett, J. Lui, D. W. Rodgers, R. L. Garlick, G. E. Tarr, Y. Husain, E. L. Reinherz, and S. C. Harrison. 1990. Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. Nature. **348**:411-418.

Watkins, B. A., H. H. Dorn, W. B. Kelly, R. C. Armstrong, B. J. Potts, F. Michaels, C. V. Kufta, and M. Dubois Dalcq. 1990. Specific tropism of HIV-1 for microglial cells in primary human brain cultures. Science. **249**:549-553.

Weniger, B. G., K. Limpakamjanarat, K. Ungchusak, S. Thanprasertsuk, K, Choopanya, S. Vanichseni, T. Uneklabh, P. Thongcharoen and C. Wasi. 1991. The epidemiology of HIV infection and AIDS in Thailand. ADIS. **5**(suppl 2): S71-S85.

Winkelstein W., and A. M. Johnson. 1990. Epidemiology overview. ADIS. **4**(suppl 1): S95-S97.

Weiss, R. A., N. N. Teich, H. E. Varmus, and J. Coffin. 1985. RNA Tumor Viruses. Cold Spring Harbor Laboratory,

Westervelt, P., H. E. Gendelman, and L. Ratner. 1991. Identification of a determinant within the human immunodeficiency virus 1 surface envelope glycoprotein critical for productive infection of primary monocytes. Proc. Natl. Acad. Sci. U. S. A. **88**:3097-3101.

Westervelt, P., D. B. Trowbridge, L. G. Epstein, B. M. Blumberg, Y. Li, B. H. Hahn, G. M. Shaw, R. W. Price, and L. Ratner. 1992. Macrophage tropism determinants of human immunodeficiency virus type 1 in vivo. J. Virol. **66**:2577-2582.

Willerford, D. M., M. J. Gale,Jr., R. E. Benveniste, E. A. Clark, and W. M. Gallatin. 1990. Simian immunodeficiency virus is restricted to a subset of blood CD4+ lymphocytes that includes memory cells. J. Immunol. **144**:3779-3783.

Willey, R. L., R. A. Rutledge, S. Dias, T. Folks, T. Theodore, C. E. Buckler, and M. A. Martin. 1986. Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome retrovirus. Proc. Natl. Acad. Sci. U. S. A. **83**:5038-5042.

Winship, P. R. 1989. An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. Nucleic. Acids. Res. **17**:1266.

Wolfs, T. F., J. J. de Jong, H. Van den Berg, J. M. Tijnagel, W. J. Krone, and J. Goudsmit. 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.

Wolfs, T. F., G. Zwart, M. Bakker, and J. Goudsmit. 1992. HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. Virology. **189**:103-110.

Wolfs, T. F., G. Zwart, M. Bakker, M. Valk, C. L. Kuiken, and J. Goudsmit. 1991. Naturally occurring mutations within HIV-1 V3 genomic RNA lead to antigenic variation dependent on a single amino acid substitution. Virology. **185**:195-205.

Wolinsky, S. M., C. M. Wike, B. T. M. Kobber, C. Hutto, W. Parks, L. L. Rosenblum, K. J. Kunstman, M. R. Furtado, and J. L. Munoz. 1992. Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. Science **255**:1134-1137.

Wong Staal, F. and R. C. Gallo. 1985. Human T-lymphotropic retroviruses. Nature. **317**:395-403.

World Health Organization Global Statistics. 1992. AIDS. **6**:1229-1233

Wright, F. 1993. Sequence Analysis course SA3: Phylogenetic trees from molecular sequences. Scottish Agricultural Statistic Service Computational Molecular Biology.

Yokoyama, S. and T. Gojobori. 1987. Molecular evolution and phylogeny of the human AIDS viruses LAV, HTLV-III, and ARV. J. Mol. Evol. **24**:330-336.

Zack, J. A., S. J. Arrigo, and I. S. Chen. 1990*a*. Control of expression and cell tropism of human immunodeficiency virus type 1. Adv. Virus. Res. **38**:125-146.

Zack, J. A., S. J. Arrigo, S. R. Weitsman, A. S. Go, A. Haislip, and I. S. Chen. 1990*b*. HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals

a labile, latent viral structure. Cell. **61**:213-222.

Zack, J. A., A. M. Haislip, P. Krogstad, and I. S. Y Chen. 1992. Incompletely Reverse-Transcribed Human Immunodeficiency Virus Type 1 Genomes in Quiescent Cells Can Function as Intermediates in the Retroviral Life Cycle. J. Virol **66**:1717-1725.

Zhang, L. Q., P. Simmonds, C. A. Ludlam, and A. J. Leigh Brown. 1991. Detection, quantification and sequencing of HIV-1 from the plasma of seropositive individuals and from factor VIII concentrates. AIDS **5**:675-681.

Zwart, G., H. Langedijk, L. van der Hoek, J. J. de Jong, T. F. Wolfs, C. Ramautarsing, M. Bakker, A. de Ronde, and J. Goudsmit. 1991. Immunodominance and antigenic variation of the principal neutralization domain of HIV-1. Virology. **181**:481-489.

APPENDIX

Contains 6 publications (3 on HIV, 3 on HCV research) on which Mr. Zhang is coauthor (1 first author).

Zhang *et al.*, (1991), AIDS 5, 675-681.

Simmonds *et al.*, (1991), Lancet 336, 1469-1472.

Simmonds *et al.*, (1991), J. Viorl. 65, 6266-6276.

Holmes *et al.*, (1992), PNAS. 89, 4835-4839.

Watson *et al.*, (1992), Brit. of Haematol. 80, 514-518.

Garson *et al.*, (1992), Vox Sang. 62, 218-223.

Hepatitis C quantification and sequencing in blood products, haemophiliacs, and drug users

PETER SIMMONDS LIN QI ZHANG HENRY G. WATSON
SELMA REBUS ELAINE D. FERGUSON PETER BALFE
GLYNIS H. LEADBETTER PENG LEE YAP JOHN F. PEUTHERER
CHRISTOPHER A. LUDLAM

The polymerase chain reaction (PCR) detected specific hepatitis C viral (HCV) RNA sequences in plasma from 15 of 21 haemophiliacs (12 HCV-antibody positive) and 7 of 27 intravenous drug users (13 HCV-antibody positive). Quantification of RNA-positive samples showed high levels of HCV (10^5 to 10^6 copies of RNA/ml) in infected patients. HCV was more frequently found in haemophiliacs infected with human immunodeficiency virus (11/11 HIV-positive and 4/10 HIV-negative patients). HCV-RNA was detected in all batches of commercially available factor VIII tested and in low concentrations in some pools of plasma donations from volunteers. Factor VIII, manufactured from volunteer donations, was uniformly negative by PCR. Phylogenetic analysis of viral sequences showed two distinct groups: one was associated with intravenous drug users and the other with haemophiliacs infected with Scottish factor VIII preparations. Both were distinct from sequences found in commercially available factor VIII.

Lancet 1990; **336**: 1469–72.

Introduction

Hepatitis C virus (HCV)[1] has been identified as an important cause of non-A, non-B (NANB) post-transfusional hepatitis.[2] Many epidemiological studies are based upon the Ortho enzyme immunoassay (EIA) but, there may be a delay of up to a year between exposure to HCV and seroconversion.[2] Furthermore, this test may give false-positive results.[3] An anti-HCV recombinant immunoblot assay (RIBA) failed to confirm the presence of specific antibody in over 70% of EIA-reactive blood donations.[4]

HCV-RNA sequences have been found in liver[5] and plasma[6] of infected individuals with the polymerase chain reaction (PCR). We now report on the relation between HCV viraemia (measured by PCR) and HCV antibody status, together with RNA quantification and sequencing, in haemophiliacs and intravenous drug users (IVDUs).

Patients and methods

Patients

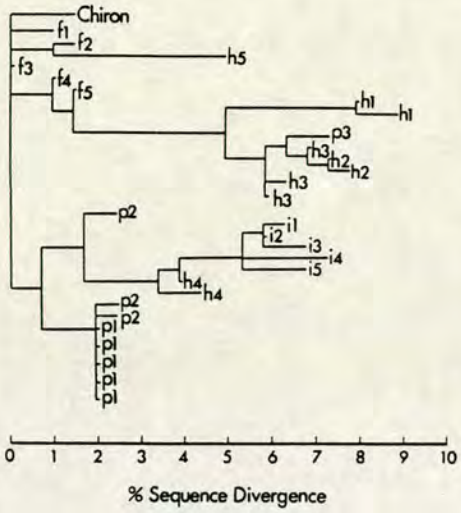
Plasma samples from 21 haemophiliacs were stored at −70°C before PCR analysis. Sera from 27 IVDUs were kept at 4°C for 3–7 days with long-term storage at −20°C before testing. Blood products were obtained from the National Institute of Biological Standards and Controls, and the Protein Fractionation Centre, Edinburgh. Freeze-dried preparations (factor VIII, factor IX) were stored at 4°C before reconstitution. Serum and plasma samples were tested by the Ortho EIA for HCV antibodies. Results are given as the optical density (OD) of the test sample divided by control OD to give the OD index (ODI). A sample was EIA-positive if ODI > 1. Most samples were further tested by the Ortho RIBA.

200 µl plasma or reconstituted blood products, 600 µl denaturing solution D, and 600 µl water-saturated phenol were mixed[7] and supplemented with 1 µg/ml purified carrier RNA (sheep fibroblast). 100 µl chloroform was added and the tube incubated on ice for 15 min. After centrifugation (14 000 *g*, 4°C, 15 min), the aqueous phase was removed and re-extracted with an equal volume of chloroform. RNA was precipitated with an equal volume of isopropanol at −20°C to −70°C for a minimum of 2 h. After further centrifugation (14 000 *g*, 4°C, 15 min), RNA was washed in 1 ml 70% ethanol solution, air-dried at 45°C, and redissolved in 20 µl water. RNA from larger samples (1 ml) was extracted by dilution with 9 ml phosphate-buffered saline, ultracentrifugation (50 000 *g*, 4°C, 3 h), and by the method described above.

RNA detection, quantification, and sequencing

Sense (ED1, GTGGTCGACTGCAATACGTGTGTCAC) and antisense (ED2, CCGGCATGCATGTCATGATGTAT) primers were used for the first reaction in a double PCR. ED3 (CACCCAGACAGTCGATTTCAG) and ED4 (GTATTTGGTGACTGGGTGCGTC) were inner (nested) primers used for the second reaction. Some samples were amplified with primers d94, d95, N1, and N2.[6] However, these primers consistently detected fewer positive samples than ED1–ED4—eg, 4/13 IVDUS were positive with d94–d95, compared with 7/13 positives with primers ED1–ED4, and were not used further. cDNA synthesis of 3 µl RNA was carried out at 42°C for 30 min with 7 units of avian myeloblastosis virus reverse transcriptase (Promega, Madison, USA) in 50 mmol/l "tris"-HCl, pH 8·0; 5 mmol/l $MgCl_2$; 5 mmol/l dithiothreitol; 50 mmol/l KCl; 0·05 µg/µl bovine serum albumin (molecular biology grade, BCL, Lewes, UK); 600 µmol/l dATP, dCTP, dGTP, and TTP; 15% dimethyl sulphoxide; 1·5 µmol/l primer ED2; 0·1 µg/µl carrier RNA (sheep fibroblast); and 10 U RNAsin (Promega). Part of the cDNA (4 µl) was amplified over twenty-five cycles with primers ED1 and ED2 in 50 µl PCR buffer.[8] 1 µl of product was amplified for a further twenty-five cycles with ED3 and ED4. Amplified DNA was detected by agarose-gel-electrophoresis and ethidium bromide staining. Quantification was by limiting dilution analysis of cDNA reverse transcribed from RNA.[8] Sequence analysis was carried out as described previously.[9] Each sequence was read in both directions, by priming with either ED3 or ED4. Phylogenetic analysis by the maximum likelihood procedure was completed with the Phylip package.[10] Sequence differences were taken as significant if the inferred G statistic exceeded the tabulated value at a probability of

ADDRESSES: **Department of Medical Microbiology, University of Edinburgh, Teviot Place, Edinburgh** (P. Simmonds, BM, S. Rebus, BSc, G. H. Leadbetter, FIMLS, J. F. Peutherer, FRCPath); **Department of Genetics, University of Edinburgh, Kings Buildings, Edinburgh** (L. Q. Zhang, MSc, P. Balfe, PhD, E. D. Ferguson, BSc); **Department of Haematology, Royal Infirmary, Lauriston Place, Edinburgh** (H. G. Watson, MRCP, C. A. Ludlam, FRCP); **and Edinburgh and South-East Scotland Regional Blood Transfusion Service, Lauriston Place, Edinburgh, UK** (P. L. Yap, MRCPath). Correspondence to Dr C. A. Ludlam, Department of Haematology, Royal Infirmary, Lauriston Place, Edinburgh EH3 9YW, UK.



Sequence diversity of HCV.

Chiron, published sequence; f 1–5, 5 commercial factor VIII batches; h 1–5, 5 haemophiliacs; i 1–5, 5 Edinburgh IVDUs; p 1–3, 3 Scottish plasma pools.

0·01. All distances indicated in the figure were significant by this method.

PCR sensitivity was assessed by equivalent experiments to those described above, but with cloned herpes simplex virus thymidine kinase (tk) gene[11] (unpublished data). An estimated 10% of cDNA transcripts extended sufficiently to be amplifiable by a pair of primers spaced 260 base-pairs apart. Equivalent experiments with known amounts of human immunodeficiency virus RNA yielded efficiencies of reverse transcription that varied with the spacing of the primers. By extrapolation we calculated an overall efficiency for detection of HCV-RNA sequences with ED1-ED4 of 5%.

Results

HCV detection and quantification

HCV antibody (EIA and RIBA) and RNA status in both IVDUs and haemophiliacs are shown in table I. HCV-RNA was more frequently detected in haemophiliacs who were also infected with human immunodeficiency virus type 1 (HIV-1). Of the 15 PCR-positive haemophiliacs, 11 were HIV-positive, whereas all 6 PCR-negative haemophiliacs were HIV seronegative. All 4 of the PCR-positive haemophiliacs who had no detectable antibody to HCV had HIV-related symptoms at the time of plasma collection, and had CD4 lymphocyte counts of $< 2 \times 10^5$/ml. The average CD4 count of the PCR-positive group was 250×10^6/l, and in the negative group 550×10^9/l ($p < 0{\cdot}05$). No association between HCV infection and either annual factor VIII use or increases in alanine aminotransferase (ALT) was found.

The amount of RNA in samples from 4 positive IVDUs and 4 haemophiliacs was estimated by titration of cDNA. PCR with nested primers detects single molecules of target DNA sequence. Tests on multiple replicates at a suitable limiting dilution give a Poisson distribution of positive and negative results that reflects the concentration of target DNA.[8] To obtain an RNA concentration from quantification of cDNA, we have assumed an overall efficiency of 5% for the reverse transcription step. The amounts of circulating RNA in 4 haemophiliacs and 4 IVDUs ranged from $3{\cdot}5 \times 10^4$ to $1{\cdot}1 \times 10^6$ copies of RNA/ml. The amounts recovered in IVDUs were similar to those of haemophiliacs. The cutoff limit of this assay was 4×10^3 copies of RNA/ml.

TABLE I—ANTIBODY (EIA AND RIBA) AND HCV-RNA STATUS IN IVDUs AND HAEMOPHILIACS

	IVDUs (n=27)		Haemophiliacs (n=21)	
—	Ab+	Ab−	Ab+	Ab−
PCR RNA+	7	0	11	4*
PCR RNA−	6	14	1	5†

*1 sample EIA-positive (ODI >5·8) but RIBA-negative.
†1 sample EIA-positive (ODI = 1·04) but RIBA-negative.

TABLE II—HCV-RNA STATUS IN BLOOD PRODUCTS

	Factor VIII							
PCR RNA	Commercial*	Scottish BTS†	UK non-commercial‡	Commercial heat-treated§	Scottish BTS heat-treated¶	Factor IX†	iv IgG†	Plasma pools**
+	10	0	0	0	0	0	0	3
−	0	4	5	1	4	5	6	4

BTS = blood transfusion service; *expiry date 1980–83; †volunteer donations 1983; ‡volunteer donations 1982–83; §wet heat-treated (60°C, 20h); ¶ dry heat-treated (80°C, 72h); and **1000 Scottish blood donations 1990.

All unheat-treated commercial batches factor VIII tested were RNA-positive (table II). Quantification by limiting dilution of two of these samples gave RNA concentrations in the original material of 2×10^4 to 10^5 copies of RNA/ml. All 9 of the non-commercial factor VIII concentrates prepared from volunteer blood donations were negative. The cutoff limit of the PCR assay for initial screening was 2000 RNA molecules/ml. To investigate whether factor VIII prepared from volunteer blood contained quantities of HCV below the threshold of sensitivity for the RNA PCR, RNA from larger volumes of Scottish concentrate was retested in an assay with a cutoff sensitivity of 200 copies of RNA/ml. All 4 batches tested remained negative. All batches of factor IX and intravenous IgG were negative for HCV-RNA. To examine whether these negative results were due to an absence of infectious blood donors, seven plasma pools that each contained 1000 volunteer donations, were tested by the PCR. Three of seven pools were positive for HCV-RNA. Samples contained 200–1900 copies of RNA/ml (test threshold 200 copies/ml). This result suggested an average frequency of < 1 PCR-positive donation/thousand in the local donor population. All plasma pools were negative for HCV antibody by the Ortho EIA (ODIs $< 0{\cdot}2$). However, as antibody titres in this test were extremely low in positive sera (ranging from only 1/10 to 1/100 in sera from 8 IVDUs) the Ortho EIA cannot detect contamination of plasma pools by low numbers (< 10) of antibody-positive donations.

Phylogenetic analysis

Nucleotide sequencing of HCV was completed by isolation of single cDNA molecules, amplification with nested primers ED1–ED4, and direct sequencing.[9] Sequences corresponded to those for a non-structural protein homologous to NS3 in flaviviruses.[12] Nucleotide substitutions were seen at 42 of 216 sites. No gaps or stop codons that would interrupt the reading frame of the nucleotide sequence were found. Almost all nucleotide changes left the encoded peptide sequences unchanged (synonymous substitutions). 17 times more synonymous nucleotide substitutions were in HCV sequences than would be expected in randomly mutating DNA,[13] which indicates a strong selection pressure on HCV against

changes in the encoded protein. Almost all substitutions were transitions, whereby a purine is substituted for another purine, or a pyrimidine is substituted by a pyrimidine.

The likely evolutionary relations between the variants of HCV are shown in the figure. The extent of differences (evolutionary distance) between variants is shown by horizontal lines. Closely related sequences are found in all IVDUs (i1-5), which suggests that they were infected from a common source. These sequences are also similar to two from an 8-year-old haemophiliac, who was first infected from factor VIII produced in Scotland in 1984 or 1985 (h4). Three haemophiliac sequences (h1–3), and an HCV sequence present in one of three positive plasma pools (p3) were also closely related, but distinct from the IVDU group. These 3 haemophiliacs may have been infected at the same time from cryoprecipitate in the mid-1960s. Sequences in two other plasma pools (p1, p2) form a third closely related group. All five batches of commercial factor VIII (f1–5) contained similar sequencies to that of the published HCV genome.[14] Finally, one sequence from an Edinburgh haemophiliac (h5) was highly divergent from the commercial factor VIII group, and from the 2–3 groups identified in Edinburgh individuals.

Discussion

Haemophiliacs had higher rates of detectable HCV infection by PCR (15/21) than by antibody tests (12/21). Of the 5 who showed no evidence of infection, 1 was a 3-year-old boy who received dry heat-treated concentrate only. The other 4 were moderate to severe haemophiliacs with high annual factor VIII use (>19 000 U/year; mean 3 years); all previously received unheated concentrates and have intermittently or persistently increased ALTs. The observation that 1 of the 4 seronegative haemophiliacs was previously antibody positive (ODI 3·2 in 1987 and 2·2 in 1988) suggests that these individuals had been infected with HCV, but had cleared the virus more rapidly than the others after the introduction of heat-treated factor VIII in 1985. The continued liver enzyme abnormalities may be due to an uncharacterised viral agent. Alternatively, as titres of antibody in the Ortho EIA are extremely low in positive samples (1/10–1/100), apparent sero-reversion may take place despite continued infection. Weiner et al[5] failed to detect HCV antibody in 2 patients with chronic posttransfusional liver disease, despite finding HCV-RNA in liver tissue.

Higher rates of HCV infection were found in haemophiliacs infected with HIV. Haemophiliacs who were PCR-positive and antibody-negative all had AIDS and low CD4 cell counts ($<200 \times 10^9$/l). Negative antibody results may be due to either this immunodeficient state or to high rates of viral protein expression that could adsorb circulating antibody by immune complex formation.

In contrast to the haemophiliacs, almost half of IVDU antibody-positive sera were negative by PCR. The use of serum rather than plasma, and the uncertain storage conditions of IVDU samples may have contributed to the failure to detect HCV-RNA. Alternatively, the viraemia may have been below the level of detection of the assay because the lower limit of the observed range was close to the assay threshold.

The finding of high rates of HCV-RNA in infected individuals is not consistent with a previous report that only 1 of 6 HCV antibody-positive blood donations transmitted infection to recipients.[6] It is possible that only individuals with high levels of circulating HCV are infectious but this is not consistent with the high rates of HCV infection associated with use of English and Scottish factor VIII that have low or undetectable levels of HCV-RNA[15] (table II). It is more likely that the 5 non-infectious, HCV–RNA donations had given false-positive results in the Ortho EIA.

The considerable sequence diversity of HCV in different patient groups resembled that of other RNA viruses with geographically separated variants—eg, the *pol* gene sequences of African and North American HIV-1.[16,17] HCV sequences were divided into three distinct groups: IVDUs (il-5), locally infected haemophiliacs (h1–3), and those who received commercial factor VIII concentrates (f1–5). The finding of related sequences in IVDUs is not surprising because needle sharing was common. The close relation between the second group of sequences obtained from five batches of factor VIII from different manufacturers is surprising in view of the wide geographical area from which paid donations are collected.

3 of 5 Edinburgh haemophiliacs were infected with a different HCV variant which was also found in one of the Scottish blood donations (p3). These 3 individuals are aged 20–30 years and were probably first infected in the 1960s with locally collected fresh frozen plasma or cryoprecipitate. These 3 individuals differ from h4, who is now aged 8 and was infected no earlier than 1984 or 1985, and from h5, who has received commercial factor VIII. The group h1–3 may consist of a variant of HCV common in Scotland 25 years ago with the IVDU/h4 sequence types becoming more prevalent in the early 1980s. However, it is not known whether reinfection with HCV can take place, and it is possible that variants in h1–3 are the results of reinfection at any time between the mid-1960s and mid-1980s.

We thank Dr G. Kemball-Cook and Dr D. P. Thomas for samples of commercial and UK non-commercial factor VIII for PCR testing; Dr R. J. Perry, Dr H. Hart, and Dr A. Todd for the supply of blood products manufactured by the Protein Fractionation Centre, Edinburgh; and Dr R. Al-Shawi for cloned tk sequences and tk primers.

REFERENCES

1. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood borne non-A, non-B viral hepatitis genome. *Science* 1989; **244:** 359–62.
2. Alter HJ, Purcell RH, Shih JW, et al. Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A, non-B hepatitis. *N Engl J Med* 1989; **321:** 1494–500.
3. MacFarlane IG, Smith HM, Johnson PJ, Bray GP, Vergani D, Williams R. Hepatitis C virus antibodies in chronic active hepatitis: pathogenic factor or false-positive result? *Lancet* 1990; **335:** 754–57.
4. Skidmore S. Recombinant immunoblot assay for hepatitis C antibody. *Lancet* 1990; **335:** 1346.
5. Weiner AJ, Kuo G, Bradley DW, et al. Detection of hepatitis C viral sequences in non-A, non-B hepatitis. *Lancet* 1990; **335:** 1–3.
6. Garson JA, Tedder RS, Briggs M, et al. Detection of hepatitis C viral sequences in blood donations by "nested" polymerase chain reaction and prediction of infectivity. *Lancet* 1990; **335:** 1419–22.
7. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 1987; **162:** 156–59.
8. Simmonds P, Balfe P, Peutherer JF, et al. Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. *J Virol* 1990; **64:** 864–72.
9. Simmonds P, Balfe P, Ludlam CA, Bishop JO, Leigh Brown AJ. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. *J Virol* 1990; **64:** 5840–50.
10. Felsenstein J. Phylogenies from molecular sequences: inference and reliability. *Ann Rev Genet* 1988; **22:** 521–65.
11. Al-Shawi R, Burke J, Jones CT, et al. A *Mup* promoter-thymidine kinase reporter gene shows relaxed tissue-specific expression and confers male sterility upon transgenic mice. *Mol Cell Biol* 1988; **8:** 4821–28.

12. Miller RH, Purcell RH. Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. *Proc Natl Acad Sci USA* 1990; **87:** 2057–61.
13. Li W-H, Wu C-I, Luo C-C. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol* 1985; **2:** 150–74.
14. Houghton M, Choo QL, Kuo G. *Eur Patent Appl* 88, 310, 992.5 and *Publ* 318, 216.
15. Garson JA, Preston FE, Makris M, et al. Detection by PCR of hepatitis C virus in factor VIII concentrates. *Lancet* 1990; **335:** 1473.
16. Alizon M, Wain-Hobson S, Montagnier L, Sonigo P. Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. *Cell* 1986; **46:** 63–74.
17. Starcich BR, Hahn BH, Shaw GM, et al. Identification and characterisation of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell* 1986; **46:** 637–48.

Effacement of glomerular foot processes in kwashiorkor

MICHAEL H. N. GOLDEN STANLEY E. H. BROOKS DAN D. RAMDATH
ELAINE TAYLOR

In a study of the pathogenesis of the oedema of kwashiorkor the ultrastructure of the kidneys from 6 children was examined shortly after they died from oedematous malnutrition. There was a generalised effacement of the glomerular epithelial cells onto the basement membrane. The filtration slits that remained were narrowed. The picture was similar to that seen in minimal-change nephrotic syndrome—but none of the children had albuminuria. The degree of effacement was statistically related to treatment with gentamicin. The findings suggest that there is a defect in the anionic charge of the glomerular basement membrane in oedematous malnutrition, that the membrane charge is more easily neutralised by cations such as gentamicin, and that, because proteinuria is not a feature of oedematous malnutrition, the proteinuria in other conditions associated with glomerular epithelial cell effacement (eg, minimal-change nephrotic syndrome) is due to something more complex than simple loss of charge.

Lancet 1990; **336:** 1472–74.

Introduction

The kwashiorkor syndrome is a generalised disorder characterised by oedema. There are changes in the cellular metabolism of sodium and potassium[1] and abnormalities in the renal handling of salt, water, acid, and osmolal loads; proteinuria, however, is uncommon.[2,3] How these abnormalities arise is unknown.[4] We have thus examined the ultrastructure of kidneys taken immediately after death from children with oedematous malnutrition.

Subjects and methods

6 children aged 4–18 months, were studied. 3 had kwashiorkor and 3 marasmic-kwashiorkor; all had nutritional oedema, fatty liver, and bacterial infection, and all had been treated according to published regimens.[5] Their plasma urea (range 1·0 to 2·8 mmol/l) and plasma creatinine (30 to 100 μmol/l) were not raised, and no patient had proteinuria. The only recognised nephrotoxin given to the children was gentamicin (5 mg/kg per day). Tissues taken within 2 h of death (table) were fixed in glutaraldehyde and processed by standard procedures. Sections from at least three blocks were stained with uranyl acetate and lead citrate.[6] Between 15 and 19 positive prints were recorded of glomeruli from each child. These prints were coded and scored blindly. A score of 0 to 4 was assigned to each photograph on the basis of whether there were no filtration slits (0), occasional narrow slits (1), moderate narrow and/or occasional normal slits (2), numerous narrow or moderate normal slits (3), or normal filtration slits (4). The scores for each picture from a subject were added up and expressed as a percentage of the maximum score achievable: a low score thus represents podocyte foot-process effacement.

The study was approved by the ethnics committee of the University of the West Indies.

Results

Case 5 had terminal acute renal failure secondary to hypovolaemia. None of the other children showed clinical evidence of renal abnormality.

Light microscopy

Light microscopy did not reveal any abnormality of the glomeruli or tubular necrosis except in case 5, in which there was some suggestion of mesangial proliferation, and in which the proximal tubules showed oedema and inclusion bodies with tubulorrhexis.

Electron microscopy

All sections showed glomerular basement membranes of uniform thickness, with clear delineation of the three laminae. No dense deposits were present in the glomerular loops or in the mesangium (fig 1). Both visceral and parietal epithelial cells were prominent. There were varying degrees of epithelial cell oedema, but vacuolation was not a feature. In all subjects the epithelial cell foot processes were effaced, (fig 1) and occurred throughout the glomeruli (fig 2). The degree of effacement varied between patients (table) but within an individual the pattern of effacement was uniform. Many of the filtration slits were narrow.

Simple regression analyses were done with height, weight, age, gentamicin total dose (per kg), time-on-gentamicin, length of time from stopping gentamicin to death, time from admission to death, admission urea, and admission creatinine as the independent variables and the podocyte effacement score as the dependent variable. Multiple regression analysis was then done with the variables most closely associated with the podocyte score. None of the variables except those associated with

ADDRESSES: **Tropical Metabolism Research Unit** (M. H. N. Golden, FRCP, D. D. Ramdath, PhD), **and Department of Pathology** (S. E. H. Brooks, FRCPath, E. Taylor, DMT), **University of the West Indies, Mona, Kingston 7, Jamaica, West Indies.** Correspondence to Dr M. H. N. Golden.

British Journal of Haematology, 1992, **80**, 514–518

Use of several second generation serological assays to determine the true prevalence of hepatitis C virus infection in haemophiliacs treated with non-virus inactivated factor VIII and IX concentrates

HENRY G. WATSON, CHRISTOPHER A. LUDLAM, SELMA REBUS,† LIN QI ZHANG,* JOHN F. PEUTHERER† AND PETER SIMMONDS† *Department of Haematology, Royal Infirmary, Edinburgh, *Institute of Animal, Cell and Population Biology, University of Edinburgh, and †Department of Medical Microbiology, University of Edinburgh*

Received 21 October 1991; accepted for publication 30 November 1991

Summary. To investigate the prevalence of hepatitis C virus infection in two risk groups, stored serum samples from treated haemophiliacs and intravenous drug users were tested for anti-HCV by both anti-C-100 based and second generation ELISAs (Abbott and Ortho) followed by testing in two confirmatory immunoblot assays that incorporate core as well as other non-structural antigens (Innogenetics LIA and Chiron RIBA-HCV test). Clear evidence of HCV infection was found in all but one of 78 haemophiliacs treated with non-virus inactivated clotting factor concentrates, but in none exposed only to super dry heat-treated concentrates. Only four samples gave rise to conflicting serological results between the four tests, two of these occurred in patients with advanced HIV related disease and almost certainly reflected loss of humoral immunity associated with disease progression, and the others occurred in the only two patients tested who were chronic carriers of hepatitis B infection and may reflect an interaction between the two viruses. Comparison of anti-C-100 versus second generation tests in immunocompetent drug users revealed a false negative rate of 20% using C-100 alone, indicating the advantage of using second generation assays for detection of past or current HCV infection. Of all of the antigens used in the confirmatory assay, positive sera showed strongest and most frequent reactivity with the C22 and C33c proteins (Ortho RIBA).

Assays for antibodies directed against the non-structural C-100 peptide, detect the majority of cases of HCV infection (Kuo *et al.* 1989). Such assays are, however, hindered by the frequent occurrence of false positive and negative results (Skidmore, 1990; Alter *et al.* 1989). It has previously been demonstrated that all recipients of non-virus inactivated factor VIII concentrates develop non-A non-B hepatitis (Fletcher *et al.* 1983); however despite this, testing with C-100 based assays has shown a prevalence of infection of between only 59% and 85% (Makris *et al.* 1990; Ludlam *et al.* 1989). Therefore using such assays, a significant proportion of haemophiliacs with a history of exposure to non-virus inactivated concentrates, many of whom have biochemical evidence of chronic liver disease, show no evidence of infection with HCV. Additionally, HCV RNA has been detected by polymerase chain reaction (PCR) in haemophiliacs who are seronegative for HCV by C-100 testing (Simmonds *et al.* 1990).

Correspondence: Dr P. Simmonds, Department of Medical Microbiology, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG.

Further problems have arisen in the field of blood donor screening in order to prevent post-transfusional non-A non-B hepatitis (PTNANBH). Studies in which the value of screening for HCV antibodies have been evaluated show a decrease in incidence, but not abolition of PTNANBH, when units positive for anti-C-100 are excluded (Esteban *et al.* 1990).

New antibody assays which use structural HCV antigens have recently become available and initial studies have shown high prevalence of HCV infection in patients with both PTNANBH and sporadic chronic active hepatitis (Marcellin *et al.* 1991; Craxi *et al.* 1991).

We have re-evaluated the serological status of a large group of haemophiliacs attending one haemophilia centre and a group of immunocompetent HIV-negative individuals with a history of intravenous drug use (IVDUs), for evidence of HCV infection. Our results demonstrate the different prevalence of HCV infection in the patient groups and

Table I. Detection of HCV antibody in different risk groups

	Abbott 1st gen. ELISA		Ortho 2nd gen. ELISA		Abbott 2nd gen. ELISA		Confirmatory assay Chiron RIBA			Confirmatory assay Innogenetics LIA		
	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Ind.	Neg.	Pos.	Ind.	Neg.
Haemophilia A												
HIV pos.	16	4	20	0	20	0	19	1	0	18	2	0
HIV neg.	36	4	39	1	39	1	38	1	1	40	0	0
Haemophilia B (HIV neg.)	16	2	18	0	18	0	18	0	0	18	0	0
H/T concentrate only*	0	7	0	7	0	7	0	0	7	0	0	7
IVDU (HIV neg.)	20	13	25	8	25	8	24	0	9	22	3	8

* Recipients of heat-treated factor VIII ($n=6$) or factor IX ($n=1$) inactivated at 80°C for 72 h (dry).

compares the performance of the different antigens used in the assays to detect anti-HCV antibodies.

METHODS AND SUBJECTS

Stored serum samples from 85 haemophiliacs attending the Edinburgh Regional Haemophilia Centre were randomly chosen for testing. The characteristics of the group were as follows: Haemophilia A ($n=66$) of which 20 were HIV-positive, 40 HIV-negative, and six were recipients of exclusively virus-inactivated concentrate. Haemophilia B ($n=19$), of which none were HIV-positive, and one had received only heat-treated products. IVDUs ($n=33$) of which all were HIV-negative both at the time of testing and in a subsequent sample at least 2 months later.

HCV testing was carried out according to the manufacturers' instruction. In the ELISAs (Abbott anti-C-100 EIA, Abbott HCV EIA second generation and Ortho HCV ELISA test system second generation) samples with an optical density over the calculated cut-off were scored as positive. Second generation assays for anti-HCV include core and further non-structural antigens in addition to C-100 coated on the solid phase. In the immunoblot assays antibody reactivity to structural and non-structural antigens contained in the Chiron RIBA-HCV (RIBA) (C-100, 5-1-1, C33c, and C22), and Innogenetics-LIA HCV (LIA) (NS-4, NS-5, C1, C2, C3 and C4), was measured. As instructed by the manufacturers, samples were considered positive if reactive with two antigens with scores of 1+ or greater, or with one band if 2+ or greater (LIA only). Indeterminate samples were those that were reactive with only one antigen (with a score of 1+ in the LIA or 1+ to 4+ in the RIBA). The frequency of positive bands was calculated in different patient groups by dividing the number of positive outcomes by the total number tested, and the average band intensity calculated by dividing the total score of all positive bands by the number of positive results. PCR using nested primers was performed on 200 μl of serum as previously described (Simmonds *et al.* 1990). Hepatitis B surface antigen (HBsAg) and core (anti-HBc) and surface (anti-HBs) antibodies were detected by radioimmunoassay. Patients with intermittent or persistent elevation of alanine aminotransferase (ALT) over the period of approximately 1 year around testing were designated as having biochemical evidence of chronic liver disease.

Table II. Reactivity of haemophiliac samples in Innogenetics and Chiron confirmatory assays

(A) Anti-C-100 (Abbott) positive samples

	Innogenetics LIA		
Chiron RIBA	Pos.	Ind.	Neg.
Pos.	67	1	0
Ind.	0	0	0
Neg.	0	0	0

(B) Anti-C-100 (Abbott) negative samples

	Innogenetics LIA		
Chiron RIBA	Pos.	Ind.	Neg.
Pos.	7	0	0
Ind.	1	1	0
Neg.	1	0	0

RESULTS

Of the 78 haemophilia A and B patients previously exposed to non-virus inactivated concentrates, 68 were positive by anti-C-100 testing (Table I). These results were true positives, as all were positive in the second generation Abbott ELISA (2nd-GAE) and Ortho ELISA (2nd-GOE), and all were confirmed in the Chiron RIBA. All but one were positive by the LIA (Table IIa). Ten samples were anti-C-100 negative. Of these, nine were positive by 2nd-GOE and 2nd-GAE, nine were positive and one indeterminate by the LIA, and seven positive, two indeterminate and one negative by the Chiron RIBA (Table IIb). Patients who had received only heat-treated concentrates were uniformly negative by all assays (Table I). Twenty of 33 IVDUs were positive by anti-C-100 testing; however, further testing in the second generation assays showed 25 positive results leaving eight negative sera and demonstrating a false negative incidence of five in 25 using the C-100 based assay alone (Table I).

Table III. Features of patients with indeterminate HCV results

ID	ALT[1]	CDC[2] status	CD[3] $\times 10^6$/l	HBV serology: Anti HB-C	Anti HB-S	HB_sAg	HCV ELISA OD: 1st GE[4]	Abbott 2nd GE[5]	Ortho 2nd GE[6]	HCV confirmation: LIA NS-4	NS-5	C_1	C_2	C_3	C_4	RIBA 5-1-1	C-100	$C33_c$	C22	HCV PCR
1	P	IV	90	+	weak +	−	0·698	>2	>2	−	−	1+	±	−	−	−	−	2+	4+	+
2	I	IV	140	weak +	weak +	−	0·377	1·733	>2	1+	−	±	±	−	−	±	±	−	4+	+
3	P	NA[3]	400	+	−	+	0·299	0·806	1·458	−	−	−	−	−	2+	±	−	−	2+	−
4	P	NA[3]	310	+	−	+	0·242	0·05	0·180	±	−	−	1+	−	1+	−	−	−	−	−

[1] ALT level over 1 year around testing P (persistent), I (intermittent) elevation.
[2] Centre for Disease Control classification of HIV related illness.
[3] Not affected.
[4] EIA OD control cut off 0·522.
[5] EIA OD control cut off 0·668.
[6] EIA OD control cut off 0·459.

Amongst the haemophiliacs, the indeterminate results were confined to four patients. The details of antibody reactivity and PCR results are given in Table III. Two of these individuals had late stage infection with HIV (CD4+ $<200 \times 10^6$/l), but were positive for HCV RNA and therefore certainly infected, the weak antibody reactivity reflecting their state of immunosuppression. The other two indeterminate samples were PCR negative, but both were antibody positive in at least one assay (Table III). Interestingly these two samples were from the only two persistent carriers of hepatitis B virus infection and may reflect an interaction between the two viruses resulting in altered expression as has been previously shown (Brotman *et al.* 1983; Tanaka *et al.* 1991).

Comparison of antibody reactivity to the individual peptides showed that in all groups C22 and C33c were the most sensitive indicators of infection with detection rates ranging 96–100% and 95·2–100% respectively (Fig 1). The intensity of antibody reactivity was in general slightly greater in HIV-negative than in HIV-positive samples (Fig 2).

DISCUSSION

Overall our results provide convincing serological evidence for the expected universal prevalence of HCV infection in haemophiliacs treated with non-virus inactivated factor VIII (Fletcher *et al.* 1983) and IX concentrates. In contrast, all seven patients exposed only to concentrates, heat treated dry for 72 h at 80°C, were uniformly negative in all serological assays (Table I). Although the numbers of patients studied was small, the absence of serological reactivity in this group provides reassurance about the viral safety of currently available Scottish National Blood Transfusion Service factor VIII and IX concentrates. Other studies have demonstrated absence of reactivity to C-100 in patients receiving only concentrate treated by this method (Evans *et al.* 1990), in contrast to others which demonstrated that the early forms of heat treatment although decreasing the incidence, did not abolish PTNANBH (Columbo *et al.* 1985; Kernoff *et al.* 1987).

Four of the 10 haemophiliacs negative for anti-C-100 were HIV-positive and therefore these false negative results may be attributable to immunosuppression as has been demonstrated with other antibody responses in HIV infected individuals (Mannucci *et al.* 1989). However false negative results in the anti-C-100 assay were also found in six haemophiliacs and five IVDUs all of whom were HIV-negative (Table I), giving an overall prevalence of false negatives of 13% in the HIV negative group. HIV negative haemophiliacs have been shown to have minor alterations in their immune systems which may reflect a degree of immunosuppression (Madhok *et al.* 1986; Carr *et al.* 1984); however, even if these are excluded, five of 25 presumed immunocompetent IVDUs still had false negative results by C-100 testing alone, and therefore it can be presumed that screening of similarly immunocompetent donors by the Blood Transfusion Service by anti-C-100 testing would fail to detect HCV infection in a considerable number of cases. Esteban *et al* (1990) have previously demonstrated that exclusion of units positive by C-100 assays does not completely abolish the risk of PTNANBH:

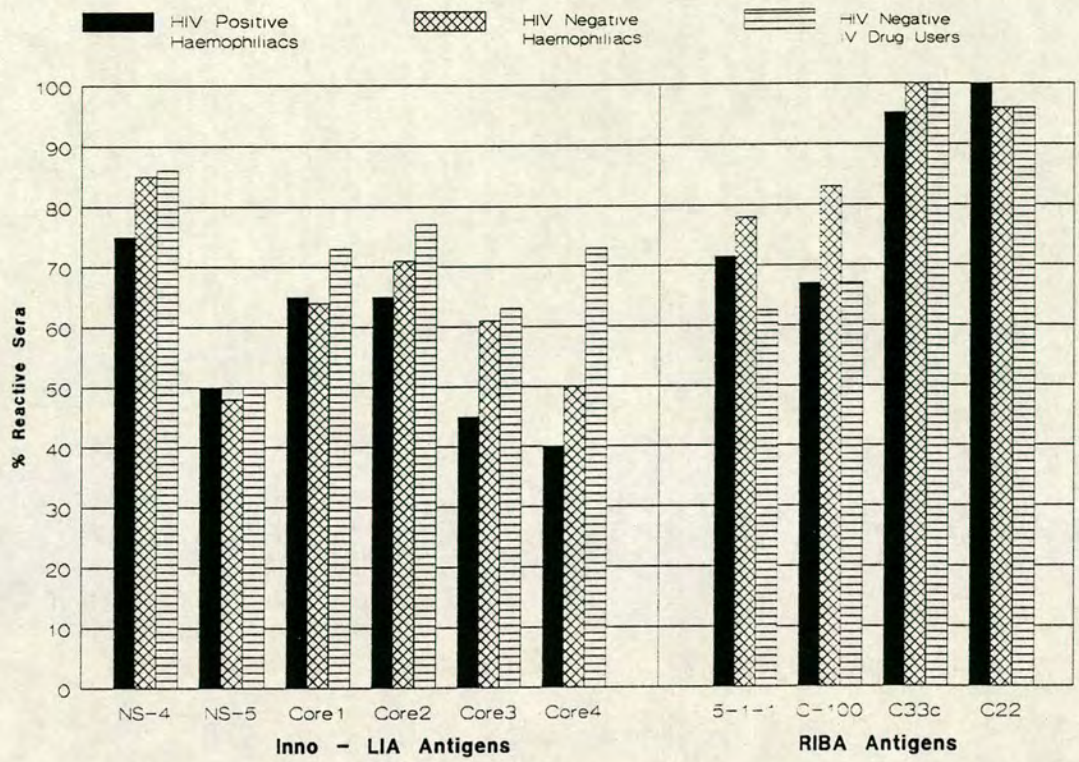


Fig 1. Frequency of positive results for individual antigens.

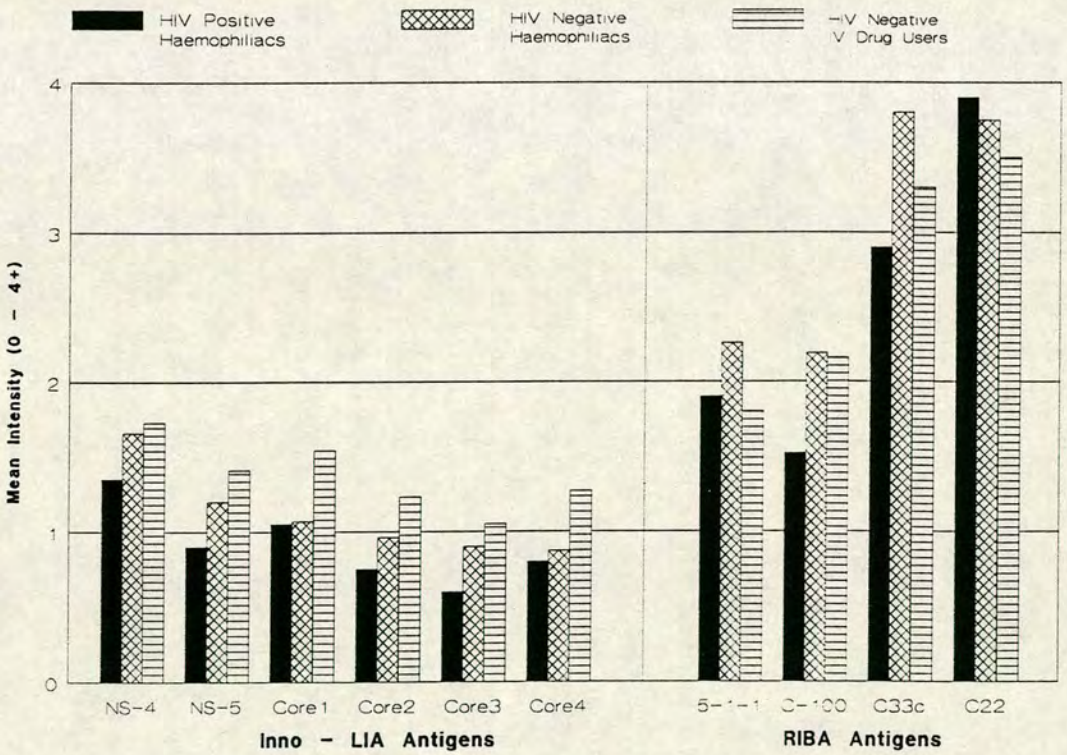


Fig 2. Mean intensity of reaction for individual antigens.

however, our data would suggest that these alternative serological assays would detect most HCV infected samples. It is, however, still possible that despite screening of donors by more sensitive second generation assays, cases of PTNANBH will occur as virus with a different serotype, not detectable by present assays may exist. In addition the long period to seroconversion may still provide a window by which infection in donors may be missed.

The finding of indeterminate, or in one case possibly negative, results of HCV serology in the only HBV surface antigen carriers is consistent with a previous study that showed a decreased prevalence of anti-HCV in patients with hepatocellular carcinoma and liver cirrhosis and who are HBsAg-positive (Tanaka *et al.* 1991), suggesting that concurrent infection interferes with viral replication (Brotman *et al* (1983).

The high frequency of positive results obtained using the antigens C22 and C33c, and the intensity of the bands seen in the immunoblot assays, suggests that these antigens are both conserved and strongly immunogenic and are the most useful of the antigens evaluated for the detection of HCV infection.

ACKNOWLEDGMENTS

The authors acknowledge the technical assistance of Mrs G. H. Leadbetter and Mr P. McCulloch in carrying out some of the serological assays. We are also indebted to n.v. Innogenetics s.a. and Ortho Diagnostics for supplying LIA strips and second generation EIA kits for evaluation. We thank Mr E. Thomson for help in producing the figures.

REFERENCES

Alter, H.J., Purcell, R.H., Shih, J.W., Melpoider, J.L., Houghton, M., Choo, Q.L. & Kuo, G. (1989) Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A, non-B hepatitis. *New England Journal of Medicine*, **321**, 1494–1500.

Brotman, B., Prince, A.M., Huima, T., Richardson, L., Van den Ende, M.C. & Pfeifer, V. (1983) Interference between non-A, non-B and hepatitis B virus infection in chimpanzees. *Journal of Medical Virology*, ii, 191–205.

Carr, R., Edmund, E., Prescott, R.J., Veitch, S.E., Peutherer, J.P., Steel, C.M. & Ludlam, C.A. (1984) Abnormalities of circulating lymphocytes subsets in haemophiliacs in an AIDS free population. *Lancet*, i, 1431–1434.

Columbo, M., Mannucci, P.M., Carnelli, V., Savidge, G.F., Gazengel, C. & Schimpf, K. (1985) Transmission of non-A, non-B hepatitis by heat-treated factor VIII concentrate. *Lancet*, ii, 1–4.

Craxi, A., Fiorentino, G., DiMarco, V., Marino, L., Magrin, S., Fabriano, C. & Pagliaro, L. (1991) Second generation tests in the diagnosis of chronic hepatitis. C. *Lancet*, **337**, 1354.

Esteban, J.I., Gonzalez, A., Hernandez, J.M., Viladomiu, L., Sanchez, C., Lopez-Talevera, J.C., Lucea, D., Martin-Vega, C., Vidal, X., Esteban, R. & Guardia, J. (1990) Evaluation of antibodies to hepatitis C virus in a study of transfusion-associated hepatitis. *New England Journal of Medicine*, **323**, 1107–1112.

Evans, J.A., Pasi, K.J. & Hill, F.G.H. (1990) Consistently normal CD4, CD8 and immunoglobulin levels in hemophilic boys only treated with a new high purity factor VIII concentrate. (Abstract). *Blood*, **76**, 420a.

Fletcher, M.L., Trowell, J.M., Craske, J., Pavier, K. & Rizza, C.R. (1983) Non-A, non-B hepatitis after transfusion of factor VIII in infrequently treated patients. *British Medical Journal*, **287**, 1754–1757.

Kernoff, P.B.A., Miller, E.J., Savidge, G.F., Machin, S.J., Dewar, M.S. & Preston, F.E. (1987) Reduced risk of non-A, non-B hepatitis after a first exposure to wet-heated factor VIII concentrate. *British Journal of Haematology*, **67**, 207–211.

Kuo, G., Choo, Q.L., Alter, H.J., Gitnick, G.L., Redeker, A.G., Purcell, R.H., Miyamura, T., Dienstag, J.L., Alter, M.J., Stevens, C.E., Tegtmeier, G.E., Bonino, F., Colombo, M., Lee, W.S., Kuo, C., Berger, K., Shuster, J.R., Overby, L.R., Bradley, D.W. & Houghton, M. (1989) An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science*, **224**, 362–364.

Ludlam, C.A., Chapman, D., Cohen, H. & Litton, P.A. (1989) Antibodies to hepatitis C virus in haemophilia. *Lancet*, ii, 560.

Madhok, R., Gracie, A., Lowe, G.D.O., Burnett, A., Froebel, K., Follet, E. & Forbes, C.D. (1986) Impaired cell mediated immunity in haemophilia in the absence of infection with human immunodeficiency virus. *British Medical Journal*, **293**, 978–980.

Makris, M., Preston, F.E., Trigger, D.R., Underwood, J.C.E., Choo, Q.L., Kuo, G. & Houghton, M. (1990) Hepatitis C antibody and chronic liver disease in haemophilia. *Lancet*, ii, 1117–1119.

Mannucci, P.M., Zanetti, A.R., Gingeri, A., Tanzi, E., Morfini, M., Messori, A., Tirindelli, M.C., de Biasi, R., Ciavarella, N. & Colombo, M. (1989) Long-term immunogenicity of a plasma-derived hepatitis B vaccine in HIV seropositive and HIV seronegative hemophiliacs. *Archives of Internal Medicine*, **149**, 1333–1337.

Marcellin, P., Martinot-Peignoux, M., Boyer, N., Pouteau, M., Aumont, P., Erlinger, S. & Benhamou, J.P. (1991) Second generation (RIBA) test in diagnosis of chronic hepatitis C. *Lancet*, **337**, 551–552.

Simmonds, P., Zhang, L.Q., Watson, H.G., Rebus, S., Ferguson, E.D., Balfe, P., Leadbetter, G.H., Yap, P.L., Peutherer, J.F. & Ludlam, C.A. (1990) Hepatitis C quantification and sequencing in blood products, haemophiliacs and drug users. *Lancet*, **336**, 1469–1472.

Skidmore, S. (1990) Recombinant immunoblot assays for hepatitis C antibody. *Lancet*, i, 1346.

Tanaka, K., Hirohata, T., Koga, S., Sugimachi, K., Kanematsu, T., Ohryohji, F., Nawata, H., Ishibashi, H., Maeda, Y., Kiyokawa, H., Tokunaga, K., Irita, Y., Takeshita, S., Arase, Y. & Nishino, N. (1991) Hepatitis C and hepatitis B in the aetiology of hepatocellular carcinoma in the Japanese population. *Cancer Research*, **51**, 2842–2847.

Original Paper

Vox Sang 1992;62:218–223

Hepatitis C Viraemia in United Kingdom Blood Donors

A Multicentre Study

J. A. Garson[a]
J. P. Clewley[b]
P. Simmonds[c]
L. Q. Zhang[d]
J. Mori[b]
C. Ring[a]
E. A. C. Follett[e]
B. C. Dow[f]
S. Martin[g]
H. Gunson[h]

[a] Division of Virology, Department of Medical Microbiology, University College and Middlesex School of Medicine, London, UK;
[b] Virus Reference Laboratory, Central Public Health Laboratory, Public Health Laboratory Service, London, UK;
[c] Department of Medical Microbiology, University of Edinburgh, UK;
[d] Department of Genetics, Institute of Cell, Animal and Population Biology, Kings Buildings, University of Edinburgh, UK;
[e] Regional Virus Laboratory, Ruchill Hospital, Glasgow, UK;
[f] West of Scotland Blood Transfusion Service, Law Hospital, Carluke, UK;
[g] Edinburgh and South East Scotland Blood Transfusion Service, Edinburgh, UK;
[h] National Directorate of the Blood Transfusion Service of England and Wales, Manchester, UK

Abstract

Of 10,633 blood donations tested in three regional blood transfusion centres with two commercial first generation screening assays for antibodies to the hepatitis C virus (HCV), 65 (0.61%) were found to be repeatedly reactive in one or both assays. Five of the 65 were confirmed positive by recombinant immunoblot assay (Ortho RIBA-2) and a further 4 were judged indeterminate. All 5 RIBA-2 positive donations and 1 of the 4 RIBA-2 indeterminates were shown to be viraemic by HCV-RNA polymerase chain reaction (PCR) assays performed at three independent reference laboratories. The remaining 56 screen test reactive donations proved negative by RIBA-2 and, with 1 exception, negative by PCR. We conclude that while first generation anti-HCV screening assays generate a high proportion of false reactions when screening low prevalence populations, results of the RIBA-2 confirmatory test correlate well with PCR findings and thus indirectly with both hepatitis C viraemia and infectivity.

Introduction

Following the introduction of blood donor screening for hepatitis B surface antigen in the 1970s, it became apparent that most residual cases of post-transfusion hepatitis were of the non-A, non-B (NANB) type. The agent responsible remained unknown until the discovery of hepatitis C virus (HCV) by Choo et al. [1] in 1989. Cloning and expression of a segment (designated C100) of the non-structural region of the HCV genome rapidly led to the development of commercially available assays for antibodies to the virus [2]. Although these 'first generation' assays have proved to be relatively reliable when used to test high seroprevalence groups, such as haemophiliacs, their specificity in the context of low seroprevalence populations has repeatedly been questioned [3–5].

Studies in the United Kingdom and the Netherlands [6, 7] have shown that the correlation between screen test

Received: June 28, 1991
Revised manuscript received: Oct. 3, 1991
Accepted: Oct. 21, 1991

Dr. J. A. Garson
Division of Virology, Department of Medical Microbiology
UCMSM, Windeyer Building, 46, Cleveland Street
London W1P 6DB (UK)

0042–9007/92/0624–0218
$2.75/0

reactivity (anti-C100) and infectivity of individual donations is poor, although better than earlier tests for 'surrogate' markers of NANB hepatitis. In contrast, polymerase chain reaction (PCR) techniques [5, 8] for the detection of circulating HCV-RNA have been shown to be capable of differentiating between infectious and non-infectious donations. However, PCR techniques are not currently suitable for mass donor screening, and doubts have been raised concerning the reliability and reproducibility of PCR results between different laboratories [9]. Although HCV-RNA can be detected by dot blot hybridization [10], in the absence of a culture system or assay for circulating HCV antigen, PCR is the only adequately sensitive means of direct viral detection.

The introduction of donor screening for anti-HCV in the UK was considered by a committee advising the Health Departments. The members included representatives of the National Blood Transfusion Service in England and Wales and of the Scottish Blood Transfusion Service. In order to obtain information regarding the prevalence of hepatitis C in the UK donor population, a prospective multicentre study, coordinated by the National Directorate of the National Blood Transfusion Service, was funded by the Department of Health (London, UK). Three regional transfusion centres tested donations with two commercially available first generation anti-HCV assays. Donations that were repeatedly reactive in either or both assays were referred to three independent reference laboratories for supplementary serology [11] and PCR analysis for the detection of viraemia. In this initial report, we present in detail the results of the PCR analysis from the three reference laboratories, together with an outline of the relevant serological findings. Full data from the first generation anti-HCV screening tests, the second generation screening tests subsequently performed on the same samples and the results of further confirmatory serology will be presented in a second report.

Materials and Methods

Specimens

Blood donations were screened at three regional blood transfusion centres (North London, Glasgow and Newcastle, UK) with Ortho and Abbott first generation anti-HCV tests, in accordance with the manufacturer's instructions. Aliquots of each of the repeatedly reactive samples were referred to three reference laboratories (Virus Reference Laboratory, VRL, at Colindale, UK; University College and Middlesex School of Medicine, UCMSM, London, UK; Regional Virus Laboratory, RVL, Glasgow, UK). PCR analysis was performed for RVL by the Department of Medical Microbiology, University of Edinburgh, UK (UE).

Confirmatory Serological Assay

The referred specimens were tested with the Chiron second generation recombinant immunoblot assay (RIBA-2, Ortho Diagnostics) [11]. This assay has four recombinant HCV antigens, 5-1-1, C100-3, C33c and C22-3, and a specificity control antigen, superoxide dismutase, SOD. According to the manufacturer's criteria, a sample was judged 'positive' if it was reactive with two or more of the HCV antigens and 'indeterminate' if it reacted with one HCV antigen only.

Viraemia Detection by PCR

In each of the three independent reference laboratories, PCR was performed 'blind' by operators who had no knowledge of the results of the confirmatory serological assay. Each laboratory used a different protocol as described below. Nucleotide numbering of primers and probes are given according to the system of Kato et al. [12], unless otherwise stated.

At UCMSM, the methods employed for extraction of viral RNA from plasma and for cDNA synthesis were exactly as previously described [13], but with random hexamers used for priming cDNA synthesis in place of anti-sense primer. Nested PCR was performed with the NS5 region primer set [5] and also with the 5' non-coding region primer set [14], as previously described.

At VRL, viral RNA was extracted from 50-µl aliquots of plasma by the method of Boom et al. [15] and reverse transcribed using murine leukaemia virus reverse transcriptase and random hexamer primers. PCR was performed with the following four sets of nested primers: (1) the NS3 primer set [16] with inner primers and probe developed at VRL [9]; (2) the NS3 primer set and probe [17] with inner primers 5'-AGA TGC GGT TTC GCG CACT (3,089, numbered according to the Chiron patent [18]) and 5'-CCA CAT TTG ATC CCA CGAT (3,476, numbered according to [18]); (3) the 5' non-coding region primers [14] with the addition of 5'-AGT GGT CTG CGG AAC CGG TGA GTA CAC CGG (130) as an internal probe, and (4) a novel set of 5' non-coding region primers, outer 5'-GCG ACA CTC CAC CAT AGA (7) and 5'CAC GGT CTA CGA GAC CTC CC (325), inner 5'-GTG AGG AAC TAC TGT CTT (35) and 5'-CGC AAG CAC CCT ATC AGG CA (297) with the same internal probe as in (3) above. The specificity of the PCR products was confirmed by oligomer hybridization [19].

At UE, the RNA-PCR was performed as described previously [8]. Primers for the 5' non-coding region [14] and new primers in NS3 were used. The sequence of the NS3 primers was based on comparative sequence data obtained in the NS3 region with primers ED1-4 [8]. The sequences are: ED5 (outer, 4,991) 5'-TCT TGA ATT TTG GGA GGG CGT CTT, ED6 (outer, 5,156) 5'-CTT CCA CAT CTG GTC CCA CGA TGG, ED7 (inner, 5,013) 5'-CAT ATA GAT GCC CAC TTC CTA TC and ED8 (inner, 5,116) 5'-CTA GCG CAC ACG GTG GCT TGG TA. The expected sizes of the PCR bands are 165 and 104 base pairs for the outer and inner products, respectively. HCV genome titres of PCR-positive plasma samples were estimated (with UE NS3 primers) by limiting dilution analysis of cDNA as described previously [8].

Table 1. PCR analysis of RIBA-2-positive and -indeterminate donations

Donation No.	Ortho RIBA-2						PCR					
	5-1-1	C100	C33	C22	SOD	Result	VRL NS3	5'	UCMSM NS5	5'	UE NS3	5'
1	R	R	R	R	NR	P	P	P	P	P	P	P
2	R	R	R	R	NR	P	P	P	P	P	P	P
3	R	R	R	R	NR	P	P	P	P	P	P	P
4	NR	R	R	R	NR	P	E	P	N	P	P	P
5	R	R	R	R	NR	P	P	P	N	P	P	P
6	NR	NR	NR	R	NR	I	N	P	N	P	P	P
7	R[1,2]	NR	NR	NR	NR	I	N	N	N	N	N	N
8	R[2]	NR	NR	NR	NR	I	N	N	N	N	N	N
9	NR	R[1,2]	NR	NR	NR	I	N	NT	N	N	N	N
10	NR	E	NR	NR	NR	N	N	N	N	N	P	N

R = Reactive; NR = non-reactive; P = positive; I = indeterminate; N = negative; E = Equivocal, band visible but less intense than manufacturer's recommended cut-off in the RIBA-2 test or an ambiguous result by PCR; NT = not tested; 5' = 5' non-coding region.

[1] Result when tested at RVL.

[2] Result when tested at VRL.

Table 2. HCV genome titres of PCR-positive donations

Donation No.	cDNA volume µl	Frequency of positives	HCV genome titre copies/ml plasma
1	1×10^{-3}	18/23	2×10^5
2	1×10^{-4}	12/23	1×10^6
3	1×10^{-6}	8/13	1×10^8
4	1×10^{-4}	5/13	5×10^5
5	1×10^{-4}	7/13	1×10^6
6	1×10^{-3}	6/23	5×10^4
10	4 µl[1]	3/3	50

cDNA volume = Amount of cDNA in replicates at the limiting dilution used for quantitation. The frequency of positives/the number tested at the limiting dilution and the estimated minimum number of HCV-RNA molecules per millilitre of plasma, calculated as described [8], are given.

[1] Positive results reproducibly obtained when 4 µl cDNA was amplified; negative at all dilutions used for titration.

Results

A total of 10,633 blood donations were tested by the three regional blood transfusion centres. Sixty-five (0.61%) were repeatedly reactive in either one or both of the commercial (anti-C100) screening assays, and these were referred to each of the three reference laboratories. Five of the 65 were confirmed positive on RIBA-2 testing, 4 gave an indeterminate result, and the remaining 56 were RIBA-2 negative (table 1).

The 65 referred samples were also tested by PCR for the presence of HCV-RNA. Six samples were found to be PCR positive by all three reference laboratories working independently and employing different experimental protocols (table 1). Five of the 6 PCR-positive samples were RIBA-2 positive, and the other was RIBA-2 indeterminate. All 6 were reactive in both the Abott and the Ortho screening assays.

Several different sets of nested PCR primers amplifying different regions of the HCV genome were used in the study. The NS3 primers used by VRL (two sets) detected 4 of the 6 positives, whereas the NS3 primer set used by UE detected all 6. The NS5 primer set used by UCMSM detected only 3. The most consistent results were generated by the NS3 primers from UE and the 5' non-coding region primers which detected all 6 of the positives at VRL, UE and UCMSM. Quantification of HCV-RNA was performed at UE by limit dilution analysis as previously described [8]. Viraemia titres of the 6 agreed PCR-positive samples ranged from 10^4 to 10^8 HCV genomes/ml of plasma (table 2).

In addition to the 6 agreed PCR positives, a further sample (donation No. 10, tables 1, 2) was found to be PCR positive with the UE NS3 primer set only. Repeat testing at UE of the UE aliquot and of a separate aliquot of the

same donation sent from UCMSM confirmed the UE NS3 primer set result. Titration experiments demonstrated that the HCV genome titre of the sample was just at the cut-off level of the UE NS3 PCR assay (approximately 50–100 HCV genomes/ml). Sequencing studies revealed that the NS3 sequence from donation No. 10 was quite distinct from that of other HCV 'isolates' present in the UE laboratory at that time. A repeat donation from the same donor given approximately 9 months later was found to be negative by RIBA-2, negative by PCR with 5′ non-coding region primers at UCMSM and UE, but positive as before with the UE NS3 primer set. Once again, the HCV genome titre was just at the cut-off level of the UE NS3 PCR assay.

Discussion

What does this study tell us about the 'true' prevalence of HCV infection in the UK blood donor population? The 0.61% (65 of 10,633) 'crude' seroprevalence observed here is consistent with a number of other reports of blood donor screening using the first generation Ortho anti-C100 ELISA [6, 7]. It is apparent however that the majority (55 of 65; 85%) of screen-reactive donations identified by this test were false since they showed no reactivity in the RIBA-2 assay and were negative by PCR. If the calculation is based exclusively on confirmed RIBA-2-positive donations, a seroprevalence of 0.047% (5 of 10,633) is obtained. Although this figure is likely to be more accurate than the 'crude' 0.65% seroprevalence it is probably an underestimate because at least some RIBA-indeterminate donations (e.g. No. 6 of table 1) may be from HCV-infected individuals [20]. Furthermore, the anti-C100-based first generation tests will miss those donors whose serum contains antibodies to HCV proteins other than C100. Antibody assays will also fail to identify those donors in the acute phase of HCV infection prior to seroconversion. The existence of infectious but anti-C100-negative donations has been reported previously [21, 22].

The pattern of RIBA-2 reactivity (C22 band only), exhibited by the PCR-positive donation No. 6, suggests that the RIBA-2 assay may not be sufficiently sensitive to confirm all genuine anti-HCV-positive donations. Antibodies against C100 were almost certainly present in this serum because it was repeatedly reactive in both the Abbott and the Ortho anti-C100 screening assays, and yet the RIBA-2 C100 band was scored as non-reactive by all three reference laboratories. Similar observations relating to inadequate sensitivity of the RIBA-2 test have been made previously [23]. RIBA-2-'indeterminate' donations, especially those with C22 only or C33 only patterns, should therefore be regarded with due caution.

The most striking finding of the present study is the clear association between RIBA-2 positivity and PCR positivity. The association implies that most if not all RIBA-2-confirmed anti-HCV-positive donors are viraemic and therefore potentially infectious. This knowledge will undoubtedly facilitate the donor counselling that will be necessitated by the introduction of anti-HCV screening. A similar association between confirmed anti-HCV positivity and viraemia has also been observed in patients with community-acquired chronic NANB hepatitis [24] and in patients with haemophilia [25]. If subsequent blood donor studies confirm that RIBA-2 positivity is invariably or almost invariably accompanied by viraemia, then the requirement for PCR analysis of blood donations should decline greatly. Further studies, funded by the Department of Health, are underway to determine whether the close association between RIBA-2 positivity and viraemia also exists when the same 10,633 donations are tested by second generation screening assays, i.e. ELISAs incorporating both structural and non-structural HCV proteins.

In view of the complexity of the technique and doubts about the feasibility of avoiding contamination, it is reassuring to note that the three independent reference laboratories, using different PCR protocols, agreed on the same 6 PCR-positive samples from a total of 65 tested. These results suggest that PCR for HCV-RNA is a reliable and reproducible diagnostic technique particularly when primers from the highly conserved 5′ non-coding region [26] are used.

However, 1 donation (No. 10) was found to be repeatedly PCR positive, on two separate aliquots, when tested with the NS3 primer set at UE, although, in agreement with VRL and UCMSM, it proved to be PCR negative with 5′ non-coding region primers. Since the titre of HCV-RNA in this sample was just at the cut-off level of the UE NS3 PCR it seems likely that minor sensitivity differences between the UE NS3 primer set and the other primer sets may have been responsible for the discordant result. Sequence analysis of this sample appears to exclude the possibility of contamination from previously or concurrently amplified PCR products in the UE laboratory. The presence of HCV-RNA (once again detected by the UE NS3 set only) in a later sample from the same donor also appears to rule out contamination. It seems therefore that this donor has a persistent very low level viraemia below the threshold of detection by PCR, except with the UE

NS3 primer set. Further samples from this donor will be investigated to look for changes in viral titre over time and for evidence of an evolving immune response. It will be interesting to see whether subsequent studies confirm that the UE NS3 primer set is consistently more sensitive than primers based on the 5' non-coding region.

Follow up studies are planned to determine whether the HCV-PCR-positive donors identified here have evidence of hepatic dysfunction or evidence of other blood borne viral infections. It is of interest to note that, with the exception of donation No. 10, the HCV-RNA titres observed in these plasma samples are similar to those previously reported in patients with chronic NANB hepatitis [24]. Attempts will be made to discover how these donors became infected with HCV and whether or not they have transmitted the infection to other family members.

We conclude that first generation anti-HCV tests generate a high proportion of false reactions when used to screen UK blood donations. At the time of writing, the first generation tests have been replaced with assays which incorporate both structural and non-structural HCV proteins, and it will be interesting to determine whether the high frequency of false reactions has been reduced. Fortunately, in this population of UK donors, the RIBA-2 test generally appears to differentiate well between true and false reactivity as judged by the results of PCR analysis for HCV-RNA. From the data presented here, it would seem prudent to regard RIBA-2-reactive donations, whether 'positive' or 'indeterminate', as likely to be viraemic and therefore capable of transmitting HCV infection.

Acknowledgements

We thank the Blood Transfusion Service of England and Wales, the Scottish Blood Transfusion Service and the Department of Health for supporting this study. We also thank the Directors of the three regional transfusion centres, Dr. M. Contreras, Dr. H. Loyd and Dr. R. Mitchell, and their staff. We are grateful to Dr. J. Barbara and Dr. S. Aloysius for coordinating the serological aspects of the continuing national study and to Prof. R. S. Tedder, M. Briggs and C. Perrons at UCMSM and the staff of the hepatitis and HIV reference laboratory at UE. Thanks are also due to Dr. P. Mortimer, F. G. Gabriel and the staff of VRL and the Communicable Disease Surveillance Centre who helped in this project and to Dr. P. L. Yap of the South East Scotland Blood Transfusion Service.

References

1 Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M: Isolation of a cDNA clone derived from a blood borne non-A, non-B viral hepatitis genome. Science 1989;244:359–362.

2 Kuo G, Choo QL, Alter HJ, Gitnick GL, Redeker AG, Purcell RH, Miyamura T, Dienstag JL, Alter MJ, Stevens CE, Tegtmeier GE, Bonino F, Colombo M, Lee WS, Kuo C, Berger K, Shuster JR, Overby LR, Bradley DW, Houghton M: An assay for circulating antibodies to a major aetiologic virus of human non-A, non-B hepatitis. Science 1989;244: 362–364.

3 Contreras M, Barbara JAJ: Screening for hepatitis C virus antibody. Lancet 1989;ii:505.

4 Flegg PJ: Ethics of screening for hepatitis C virus. Lancet 1989;ii:1221.

5 Garson JA, Tedder RS, Briggs M, Tuke P, Glazebrook JA, Trute A, Parker D, Barbara JAJ, Contreras M, Aloysius S: Detection of hepatitis C viral sequences in blood donations by 'nested' polymerase chain reaction and prediction of infectivity. Lancet 1990;335:1419–1422.

6 Van der Poel CL, Reesink HW, Schaasberg W, Leentvaar-Kuypers A, Bakker E, Exel-Oehlers PJ, Lelie PN: Infectivity of blood seropositive for hepatitis C virus antibodies. Lancet 1990;335:558–560.

7 Contreras M, Barbara JAJ, Anderson CC, Ranasinghe E, Moore C, Brennan MT, Howell DR, Aloysius S, Yardumian A: Low incidence of non-A, non-B post-transfusion hepatitis in London confirmed by hepatitis C virus serology. Lancet 1991;337:753–757.

8 Simmonds P, Zhang LQ, Watson HG, Rebus S, Ferguson ED, Balfe P, Leadbetter GH, Yap PL, Peutherer JF, Ludlam CA: Hepatitis C quantification and sequencing in blood products, haemophiliacs, and drug users. Lancet 1990;336:1469–1472.

9 Clewley JP: Detection of hepatitis C virus RNA in serum. Lancet 1990;336:309–310.

10 Maeno M, Kaminaka K, Sugimoto H, Esumi M, Hayashi N, Komatsu K, Abe K, Sekiguchi S, Yano M, Mizuno K, Shikata T: A cDNA clone closely associated with non-A, non-B hepatitis. Nucleic Acids Res 1990;18:2685–2689.

11 Van der Poel CL, Cuypers HTM, Reesink HW, Weiner AJ, Quan S, Di Nello R, Van Boven JJP, Winkel I, Mulder-Folkerts D, Exel-Oehlers PJ, Schassberg W, Leentvaar-Kuypers A, Polito A, Houghton M, Lelie PN: Confirmation of hepatitis C virus infection by new four-antigen recombinant immunoblot assay. Lancet 1991;337:317–319.

12 Kato N, Hijikata M, Ootsuyama Y, Nakagawa M, Ohkoshi S, Sugimura T, Shimotohno K: Molecular cloning of a human hepatitis C genome from Japanese patients with non-A, non-B hepatitis. Proc Natl Acad Sci USA 1990;87:9524–9528.

13 Garson JA, Tuke PW, Makris M, Briggs M, Machin SJ, Preston FE, Tedder RS: Demonstration of viraemia patterns in haemophiliacs treated with hepatitis C virus contaminated factor VIII concentrates. Lancet 1990;336: 1022–1025.

14 Garson JA, Ring C, Tuke P, Tedder RS: Enhanced detection by PCR of hepatitis C virus RNA. Lancet 1990;336:878–879.

15 Boom R, Sol CJA, Salimans MM, Jansen CL, Wertheim van Dillen PME, Van der Noordaa J: Rapid and simple method for purification of nucleic acids. J Clin Microbiol 1990;28:495–503.

16 Weiner AJ, Kuo G, Bradley DW, Bonino F, Saracco G, Lee C, Rosenblatt J, Choo QL, Houghton M: Detection of hepatitis C viral sequences in non-A, non-B hepatitis. Lancet 1990;335:1–3.

17 Kubo Y, Takeuchi K, Boonmar S, Katayama T, Choo QL, Kuo G, Weiner AJ, Bradley DW, Houghton M, Saito I, Miyamura T: A cDNA fragment of hepatitis C virus isolated from an implicated donor of post-transfusion non-A, non-B hepatitis in Japan. Nucleic Acids Res 1989;17:10367–10372.

18 Chiron Corporation: European Patent Application 88310922.5.

19 Kellogg DE, Kwok S: Detection of human immunodeficiency virus; in Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds): PCR Protocols: A Guide to Methods and Applications. San Diego, Academic Press, 1990, pp 337–347.

20 Weiner AJ, Truett MA, Rosenblatt J, Han J, Quan S, Polito AJ, Kuo G, Choo QL, Houghton M, Agius C, Page E, Nelles MJ: HCV testing in low-risk population. Lancet 1990;336: 695.

21 Esteban JI, Gonzalez A, Hernandez JM, Viladomiu L, Sanchez C, Lopez-Talavera JC, Lucea D, Martin-Vega C, Vidal X, Esteban R, Guardia J: Evaluation of antibodies to hepatitis C virus in a study of transfusion-associated hepatitis. N Engl J Med 1990;323:1107–1112.

22 Inaba S, Fukuda M, Okochi K, Irita Y, Tokunaga K, Kiyokawa H, Maeda Y: HCV transmission after receiving anti-c100-negative blood units. Lancet 1991;337:1354.

23 Craxi A, Fiorentino G, Di Marco V, Marino L, Magrin S, Fabriano C, Pagliaro L: Second generation tests in diagnosis of chronic hepatitis C. Lancet 1991;337:1354.

24 Brillanti S, Garson JA, Tuke PW, Ring C, Briggs M, Masci C, Miglioli M, Barbara L, Tedder RS: The effect of alpha interferon therapy on hepatitis C viraemia in community acquired chronic non-A, non-B hepatitis: A quantitative PCR study. J Med Virol 1991;34: 136–141.

25 Tedder RS, Briggs M, Ring C, Tuke PW, Jones P, Savidge GF, Rodgers B, Garson JA: Hepatitis C antibody profile and viraemia prevalence in adults with severe haemophilia. Br J Haematol 1991;79:512–515.

26 Han JH, Shyamala V, Richman KH, Brauer MJ, Irvine B, Urdea MS, Tekamp-Olson P, Kuo G, Choo QL, Houghton M: Characterization of the terminal regions of hepatitis C viral RNA: Identification of conserved sequences in the 5′ untranslated region and poly (A) tails at the 3′ end. Proc Natl Acad Sci USA 1991;88:1711–1715.

AIDS

Reprinted from Volume 5 1991

CS
CURRENT
SCIENCE

Detection, quantification and sequencing of HIV-1 from the plasma of seropositive individuals and from factor VIII concentrates

Lin Qi Zhang, Peter Simmonds, Christopher A. Ludlam* and Andrew J. Leigh Brown

A highly sensitive and reliable RNA polymerase chain reaction method has been developed which has been used to detect, quantify and sequence cell-free HIV RNA directly from the plasma of seropositive individuals. Plasma from 10 out of 12 haemophiliacs tested was found to contain detectable levels of HIV-1 RNA [log mean value: 1.2×10^3 copies for Centers for Disease Control (CDC) group II patients, 5.5×10^3 copies for CDC group IV patients]. The presence of cell-free circulating virus in both symptomatic and asymptomatic individuals suggests that viral replication continues throughout the course of infection. The same procedure has been used to detect and sequence HIV-1 RNA in two batches of unheated commercial factor VIII concentrate distributed in 1981 and 1983. The sequences obtained revealed a closer relationship to North American than to African strains of HIV-1.

AIDS 1991, **5**:675–681

Keywords: HIV detection, polymerase chain reaction, plasma viraemia, factor VIII, quantification, nucleotide sequence.

Introduction

Using the polymerase chain reaction (PCR), HIV provirus can be not only detected [1–4], but also accurately quantified directly in patient's peripheral blood mononuclear cells (PBMC) [5,6]. However, detection of HIV DNA in PBMC does not indicate whether such cells are expressing viral RNA sequences or whether free virus is present in plasma or other body fluids. Several investigators have coupled a reverse transcriptase (RT) reaction step to the PCR (RNA PCR) [7–9] and have successfully detected HIV RNA both in cultured HIV-infected cell lines and in PBMC from seropositive subjects. Unfortunately these reports have not included an assessment of the sensitivity of the methods used and therefore have not determined absolute quantities of RNA.

Using the nested PCR method, we have developed a highly sensitive and quantitive RNA PCR assay. After reverse transcription, complementary (c) DNA was amplified in two sequential PCRs. As the nested PCR can detect single molecules of target DNA sequence, quantitation of HIV-specific cDNA, and by implication of HIV RNA sequences present in the original sample, can be achieved by limiting dilution as described previously for provirus quantification in PBMC [5]. The efficiency of the RT reaction was estimated by measurement of the yield of cDNA from known amounts of specific RNA sequences after reverse transcription. We have used the methods to estimate the plasma virus load in a group of HIV-infected individuals including both symptomatic and asymptomatic patients.

We have also investigated the presence of HIV-1 RNA sequences in eight batches of unheated factor VIII concentrate distributed between 1981 and 1984. HIV-1 RNA was detected in two batches of commercial factor VIII concentrate distributed in 1981 and 1983. The identity of the RNA was confirmed by nucleotide se-

From the Division of Biological Sciences, University of Edinburgh and the *Department of Haematology, Royal Infirmary, Edinburgh, Scotland, UK.

Sponsorship: Supported by the British Council and the Government of the PRC (L.Q.Z.) and the Medical Research Council AIDS Directed Programme.

Requests for reprints to: Dr Andrew J. Leigh Brown, Division of Biological Sciences, University of Edinburgh, ICAPB Genetics Building, West Mains Road, Edinburgh EH9 3JN, Scotland, UK.

Date of receipt: 13 February 1991; revised: 4 March 1991.

quencing the PCR product. Sequences obtained in the *pol* and *env* regions from these concentrates provide some information on the geographical origins of the infected blood donors.

Materials and methods

Plasma

Plasma samples were obtained from 12 HIV-1-infected haemophiliacs who had been exposed to HIV-contaminated factor VIII early in 1984 [10]. All 12 individuals seroconverted for antibody 3–10 months after receiving the factor VIII [10,11]. Patients 56, 70, 82, 83 and 84 have been classified as stage II according to Centers for Disease Control (CDC) clasification, while patients 72, 74, 77, 79, 95 and 87 are in CDC group IV and have been suffering from a range of opportunistic infections and constitutional symptoms of HIV infection. Patient 28 died in 1988. Apart from patient 72, all CDC group IV patients but none of the CDC group II patients have been receiving antiviral treatment (zidovudine). Three samples from a seronegative individual, who has no known risk factors for HIV infection, were used as negative controls. Plasma samples and cell culture supernatant were assayed for HIV-1 p24 core antigen with a commercial enzyme immunoassay (Du Pont, Stevenage, Herts, UK) according to the manufacturer's instructions. This assay was capable of detecting 7 pg/ml p24 antigen.

Factor VIII concentrates

Eight batches of factor VIII concentrate, including both commercial and National Health Service-produced material, were obtained from the National Institute of Biological Standards (Potters Bar, UK). All batches were unheated and prepared before the introduction of donor screening for anti-HIV antibodies. They were distributed in the UK between 1981 and 1984.

HIV and *tk* primers

HIV primers were synthesized by the Oswel DNA Service, Department of Chemistry, University of Edinburgh, and purified by high-pessure liquid chromatography (HPLC). The primers were based on the consensus of several published HIV sequences (HIV_{BRU}; HIV_{ELI}; HIV_{IIIB}; HIV_{MAL}; HIV_{MN} and HIV_{RF}). The primer sites were chosen for greatest conservation between these published sequences. No more than one mismatch with any of the above published HIV sequences was permitted, nor was any mismatch near the 3′ terminus. The primers for herpes simplex virus (HSV) *tk* gene were provided by R. Al-Shawi (Centre for Genome Research, University of Edinburgh, Edinburgh, UK). The sequences of the primers for both HIV template and transcribed HSV *tk* template are given here and the coordinates listed in the brackets of HIV primers are from HIV_{HXB2} (+, sense; −, antisense). HIV *pol* primers: (a) 5′-CCCAAAAGTTAAACAATGGCC (+, 2602), (b) 5′-AGAAATTTGTACAGAGATGG (+, 2653), (c) 5′-CCATTTATCAGGATGGAGTTC (−, 3245), (d) 5′-GCTGTCTTTTTCTGGCAGCAC (−, 3281); V4-V5 primers: (e) 5′-TCAGGAGGGGACCCCAGAAATT (+, 7316), (f) 5′-GGGGAATTTTTCTACTGTAAT (+, 7360), (g) 5′ CTTCTCCAATTGTCCCTCATA (−, 7665), (h) 5′-CCATAGTGCTTCCTGCTGCT (−, 7814). HSV *tk* primers: 5′-GCCAGTAAGTCATCGGCTCGGG (+), 5′-CCATCAACACGCGTCTGCGTTCG (−).

Viral RNA extraction

Five hundred microlitres of patient plasma or 500 µl HIV-infected C8166 culture supernatant was mixed with 8.5 ml phosphate-buffered saline (PBS); alternatively, 3 ml factor VIII reconstituted with the recommended volume of water was mixed with 6 ml PBS. In both cases, virus was pelleted at 45 000 **g** in a swing out rotor (Sorvall SH80, Du Pont) at 4°C for 2 h. The pellet was resuspended in 1.2 ml of a denaturing solution [2 mol/lguanidinium thiocynate, 12.5 mmol/l sodium citrate (pH 7.0), 0.25% sarcosyl, 0.05 mol/l2-mercaptoethanol, 50% v/v water-saturated distilled phenol] and RNA purification was continued as described [12].

cDNA synthesis

RNA was incubated with ribonuclease (RNase)-free deoxyribonuclease(DNase;Boehringer-Mannheim,Lewes, UK) at 37°C for 20 min in a 10 µl volume of DNase reaction buffer [50 mmol/l Tris-Cl pH 7.5, 10 mmol/l $MgCl_2$, 4 mmol/l DTT, 10 units RNAsin, 1 µg carrier RNA (sheep fibroblast cell total RNA) and 15 units RNase-free DNase]. The sample was then incubated at 80°C for 10 min to terminate the reaction. cDNA synthesis was carried out by adding an equal volume of RT reaction buffer [50 mmol/l Tris-Cl pH 8.0; 5 mmol/l $MgCl_2$; 5 mmol/l DTT; 50 mmol/l KCl; 0.05 µg/µl bovine serum albumin (BSA); 600 µmol/l each of dATP, dGTP, dTTP, dCTP; 20% dimethyl sulphoxide (DMSO); 1.5 µmol/l outer antisense primer (d or h); 10 units RNAsin (Promega, Southampton, UK) and 10 units AMV reverse-transcriptase (Promega) to the digested HIV RNA sample and incubating at 42°C for 30 min.

Measurement of RT reaction efficiency using double PCR method

The HIV RNA for cDNA synthesis was serially diluted in twofold steps, slot-blotted and quantified by hybridization with HIVBH10R.3 plasmid in comparison with a dilution series of known amounts of HIVBH10R.3 DNA; 50 ng HIVBH10R.3 plasmid DNA was used to make an HIV-specific probe using the Pharmacia Oligolabelling Kit (Milton Keynes, UK). Blot hybridization was carried out following the method of

Church and Gilbert [13]. cDNA samples were diluted, amplified in the PCR with products sequenced as described previously [5].

Measurement of RT reaction efficiency by plasmid RNA transcription

Construct pSV2gpt, containing a mouse promoter region and coding region of the HSV type 1 thymidine kinase gene (HSV *tk*) was obtained from R. Al-Shawi [15]. RNA was transcribed *in vitro* from 100 ng of the construct plasma DNA at 37°C for 1 h (20 µl volumes of 4 mmol/l Tris-Cl pH 8.0; 8 mmol/l $MgCl_2$; 2 mmol/l spermidine; 50 mmol/l NaCl; 0.01 mol/l DTT; 0.4 mmol/l each of rATP, rCTP and rGTP; 30 units of RNAsin; 100 ng/µl BSA; 10 units of T7 RNA polymerase). The concentration of pSV2gpt RNA transcripts used for subsequent cDNA synthesis and HSV *tk* plasmid DNA used for quantitative comparison was estimated by spectrophotometry at 260 nm. Twofold serial titrations of *tk* cDNA after reverse transcription with antisense primer and of HSV *tk* plasmid DNA were made prior to PCR amplification with HSV *tk* specific primers (spacing = 260 bp). Twenty-five cycles were employed and the product of PCR was analysed by agarose gel electrophoresis and ethidium bromide staining. The amount of cDNA was estimated by reference to a dilution series of HSV *tk* DNA after amplification with the same primers. This was then compared with the number of RNA sequences from which the cDNA was made.

Controls for PCR

In order to remove any possible HIV DNA contamination, either from plasmid itself [18] or from other sources, all HIV RNA samples were treated with RNase-free DNase (Boehringer-Mannheim) prior to reverse transcription. All experiments included appropriate negative controls. In order to confirm that carrier RNA itself did not generate results in the experiment, carrier RNA was subjected to double PCR with V4-V5 primers and *pol* primers. None of the samples was positive in any assay. HIV RNA samples amplified without RT also gave negative double PCR results (data not shown) demonstrating that the positives obtained were from HIV cDNA, not from HIV DNA contamination.

Results

RT reaction efficiency from plasmid RNA transcription

A known amount of RNA transcript from a cloned *tk* gene was reverse transcribed, and the cDNA titrated prior to amplification in the PCR with *tk*-specific primers. This was compared with the results of amplifying a dilution series of cloned *tk* DNA. The cut-off point, below which amplified DNA was not detectable by agarose gel electrophoresis and ethidium bromide staining, was 5.12 fg for the *tk* plasmid. Using the known molecular mass of the *tk* plasmid DNA (7270 bp × 660 g/mol per bp), this figure corresponded to 630 molecules of *tk* plasmid, or 1260 copies of target sequence (630 × 2 for double-stranded DNA). cDNA synthesized from 8 ng of HSV *tk* transcript was serially diluted prior to amplification. A dilution containing cDNA synthesized from 16 fg RNA gave a positive result and the next dilution containing cDNA from 8 fg was negative. Using the estimated molecular mass of the RNA transcript (2400 bases × 330 g/mol per base), the minimum detectable amount of cDNA corresponded to an input of 1.2×10^4 copies of RNA. The efficiency of the RT reaction, in terms of the number of molecules of amplifiable cDNA synthesized from the RNA template, is the ratio of the two figures obtained above (1260/12 000), or approximately 10%.

RT reaction efficiency using HIV RNA

Six HIV-1 RNA samples were extracted from culture supernatant of C8166 cells infected with HIV_{RF}, HIV_{IIIB} and HIV_{HXB2} The viral RNA was slot-blotted and quantified by hybridization with HIVBH10R.3 plasmid probe, in comparison with a dilution series of known amounts of HIVBH10R.3 DNA; 19.8, 9.9, 7 and 0.6 pg HIV_{RF} viral RNA, 3.0 and 1.5 pg HIV_{IIIB} and HIV_{HXB2}, respectively, were then reverse transcribed with an HIV V4-V5 outer antisense primer (h). The number of copies of cDNA in each sample after reverse transcription was estimated by limiting dilution and amplification with nested primers. The frequency of positive reactions at limiting dilution was used to calculate the molecular concentration of cDNA using a Poisson correction for positive reactions which contain more than one template molecule. [The mean number of template molecules per reaction (m) is equal to $-\ln(f_o)$, where f_o is the fraction of negative reactions.] The molecular mass of one copy of HIV-1 RNA was 6.5 ag [5] from which the RT reaction efficiency was calculated. In each case, 20 (or 24) replicates were tested at limiting dilution, and the estimates were reasonably accurate. For the first sample for instance, at a 1 in 78 000 dilution, there were two PCR positives out of 24 replicates. Using the Poisson formula to correct for multiple positives, the total number of molecules of HIV RNA detected at this dilution was estimated to be 2.1. As the molecular mass of single HIV RNA was 6.5 ag, the total amount of cDNA synthesized by the reverse transcription reaction was therefore equivalent to 1.06 pg ($2.1 \times 6.5 \times 10^{-6} \times 78\,000$). The ratio of the amount of HIV RNA reverse transcribed and the initial amount of HIV RNA available for reverse transcription gave the RT efficiency (1.06/19.8 = 5.3%). We show the results

of dilution and distribution of six independent cDNA reactions after double PCR amplification in Table 1, along with the calculated RT efficiencies.

Quantification of HIV virus particle in the plasma or serum of seropositive individuals

Plasma from 10 out of 12 haemophiliacs contained detectable levels of HIV-1 RNA. The threshold of detection can be calculated to be 200 copies of RNA per ml plasma, based on the volume of plasma from which RNA was extracted, the input volume of the reverse transcription reaction, the efficiency of reverse transcription and the proportion of cDNA used in the first PCR reaction. The estimated concentration of virus particles in plasma ranged from 1×10^3 to 3×10^4 copies per ml in the positive samples (Table 2), with a logarithmic mean value of 1.2×10^3 copies for CDC group II patients, and 5.5×10^3 copies for CDC group IV patients. There was a significant positive association between the concentration of viral RNA in plasma and the proviral abundance in PBMC DNA. The correlation coefficient between the log-transformed RNA and DNA estimates is 0.74 ($0.01 < P < 0.02$), indicating that over 50% of the variance in DNA proviral abundance can be explained by the RNA concentrations. Thus the patients with the lowest proviral abundance (83 and 84 with an average of one provirus in 14 000 and 10 000 PBMC, respectively) had less than 200 copies of RNA per ml in their plasma. On the other hand in patients 82 and 87, relatively higher concentrations of virus (8.5×10^3 and 3.0×10^4 per ml plasma) were associated with high frequencies of provirus-bearing PBMC (one in 700 and one in 589 PBMC infected, respectively). However, no correlation was found, regardless of the stage of infection, between the amount of virus in plasma and the level of p24 antigen. Five patients who were negative for p24 antigen (70, 72, 74, 77 and 95) contained over 1.0×10^3 virus particles per ml plasma. Further, no correlation between CD4+ lymphocyte depletion and amount of circulating virus was found in these individuals. For example, relatively normal CD4 counts were found in p82 despite containing 8.5×10^3 copies of viral RNA per ml. Conversely, p84 had low CD4 counts (0.05×10^9/l) yet no detectable circulating virus. Zidovudine treatment appeared to have little effect on the levels of circulating virus. Those on long-term treatment (74, 77 and 79) contained comparable levels

Table 1. Titration of HIV cDNA after reverse transcriptase (RT) reaction by dilution, distribution and double polymerase chain reaction (PCR) amplification.

Sample (pg)	Amount of RNA in RT reaction (pg)	Dilution factor of cDNA (no. positive cells/no. tested) 125	620	3100	15600	78000	390000	Calculated amount of cDNA (pg)	% RT efficiency
RF1	19.8	ND	4/4	4/4	3/4	2/24	0/4	1.06	5.0
RF2	9.9	4/4	4/4	4/4	4/24	0/4	0/4	0.44	4.0
RF3	7	4/4	4/4	4/4	4/24	0/4	0/4	0.44	6.0
RF4	0.6	3/4	5/24	0/4	0/4	0/4	0/4	0.02	3.3
IIIB	3.0	4/4	3/4	1/4	2/20	0/4	ND	0.22	7.0
HXB2	1.5	2/4	2/4	4/20	0/4	0/4	ND	0.09	6.0

ND, not done.

Table 2. Comparison of the amount of circulating virus from patients' plasma with the amount of provirus in peripheral blood mononuclear cells (PBMC), plasma p24 antigen concentration, CD4+ counts, and clinical status.

Patient	Virions per ml in plasma	Number of cells per single provirus	p24 antigen (pg/ml)	CD4+ lymphocytes ($\times 10^9$/l)	Zidovudine treatment (months)	Disease stage (CDC)
p83	$<10^2$	14000	–	0.27	–	II
p84	$<10^2$	10000	–	0.05	–	II
p77	1.26×10^3	2500	–	0.07	14	IVA
p56	1.29×10^3	ND	15	0.51	–	II
p74	3.49×10^3	2000	–	0.38	10	IVC2
p95	3.50×10^3	455	–	0.06	–	IV
p28	3.97×10^3	2718	300	0.09	–	IV
p79	4.39×10^3	3300	63	0.21	15	IVC2
p70	6.12×10^3	ND	–	0.39	–	II
p82	8.53×10^3	700	53	0.65	–	II
p72	2.76×10^4	2720	–	0.33	–	IVC
p87	2.96×10^4	589	20	0.05	17	IVC/E

–, less than 7 pg/ml; ND, not done; CDC, Centers for Disease Control.

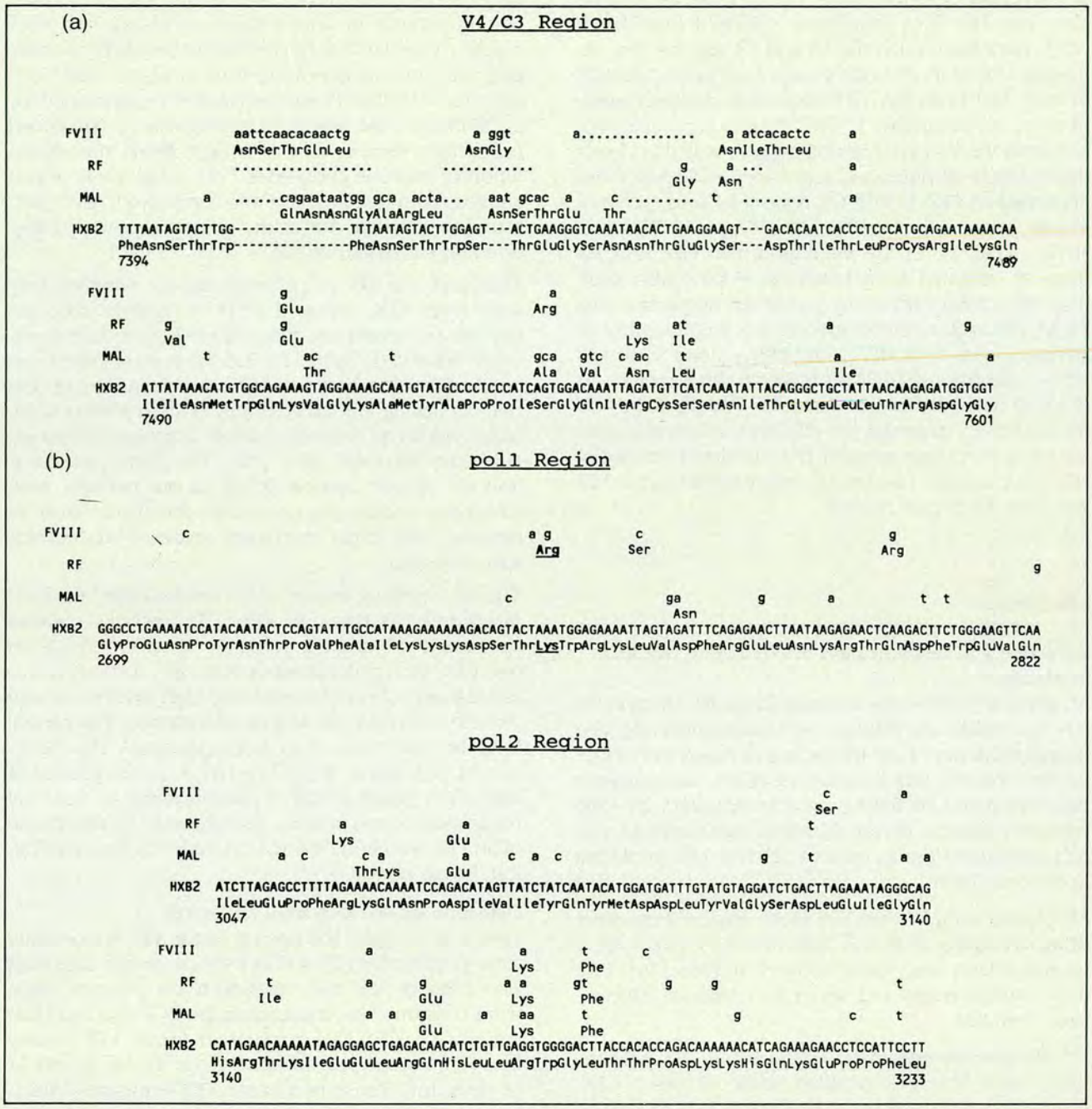


Fig. 1. Comparison of sequences detected in factor VIII with those of known geographical variants of HIV-1. (a) Nucleotide and amino acid sequences of the V4 and C3 region (*env*) of RNA detected in factor VIII batch no. 1. (b) Sequences in two regions of *pol* of RNA from batch no. 8. The location of these sequences in the genome of the HIV_{IIIB} isolate (clone HXB2, Genbank accession number K03455) is indicated. Differences between the factor VIII sequences and those of the HIV_{RF} and HIV_{MAL} isolates from that of HIV_{IIIB} are shown in the body of the figure.

to the two untreated symptomatic individuals (95 and 72).

Detection of HIV-1 in factor VIII concentrate

RNA was prepared from eight different batches of factor VIII (distributed between 1981 and 1984), by high-speed centrifugation and solubilization of the pellets as described previously. One-third of each RNA sample was reverse transcribed with primers d and h (*pol* and *env* genes), and one-quarter of the cDNA was amplified by PCR with primers a–d or e–h (see Methods). Two factor VIII batches (both commercially derived) out of eight tested gave positive results: in one case with the *env* primers, the other with the *pol* primers. Single molecules of target DNA were isolated by limit-

ing dilution of the cDNA and were directly sequenced as previously described (Fig. 1) [5,15]. In the *env* region, two HIV RNA sequences, obtained from batch no. 1, were identical in the V4 and C3 regions. The sequence was distinct from those of all published HIV isolates and from any HIV sequence obtained previously in our laboratory [15,16]. This is particularly apparent in the V4 hypervariable region, which is clearly distinct from all published sequences, including those illustrated in Fig. 1. The C3 region showed 92% homology with HIV_{HXB2}, 90% with HIV_{RF}, and 77% with HIV_{SF-2} (Fig. 1). In the *pol* region, two HIV RNA sequences obtained from batch no. 8 were also identical and distinct from any published sequence (Fig. 1). In this region, the sequence was 96% identical to HIV_{HBX2}, 94% with HIV_{RF} and HIV_{SF-2} and 92% with HIV_{Z6}. The amount of RNA present in this material was close to the threshold of sensitivity for the RNA PCR method used. Allowing 5% efficiency of reverse transcription with these primers, the calculated amount of HIV RNA in both batches of reconstituted factor VIII was only 2.5 copies per ml.

Discussion

Detection and quantification of circulating HIV RNA in plasma

An efficiency of 5% was obtained in the RT reaction for HIV-1 template and primers *e–h* (spacing 480 bp). The overall efficiency of the procedure declined with wider primer spacing; amplification of cDNA using primer pairs separated by 858 bp gave an efficiency of 1.8% (data not shown). We are therefore able to obtain a direct estimate of the amount of cell-free HIV in plasma in absolute terms.

All plasma samples used in these experiments were obtained during 1988 and 1989 from HIV-seropositive haemophiliacs who were infected in 1984 [10]. Five were asymptomatic and seven had AIDS or AIDS-related complex.

Of the five asymptomatic (CDC group II) patients, three were RNA PCR-positive, while all seven CDC group IV patients were positive for RNA PCR (Table 2). All samples that were p24 antigen-positive were positive in the RNA PCR. High levels of viral RNA sequences were found even in some p24 antigen-negative plasma samples (patients 70, 72, 74 and 77) while similar or lower amounts of circulating RNA have been found in other plasma samples that were antigen-positive. The RNA PCR method provides a direct way to detect and quantify virus production regardless of immune complex formation, hence it may provide a better marker of the progression of disease.

On average, HIV RNA was more abundant in the plasma of patients with more advanced disease compared with asymptomatics (Table 2). However, a wide range in the amount of cell-free HIV RNA was found among patients in similar stages of disease. In other studies, a similar lack of correlation between p24 antigen and titers of infectious virus in plasma has been reported [19,20]. These results can be explained by a variation in the level of p24 antibody as concluded following a recent study of antigen levels after dissolution of immune complexes [21]. Total levels of p24 antigen, both free and immune complexed, have been found to vary little during the course of primary infection and subsequently [21].

The levels of HIV in patients' plasma, whether they were from CDC group II or from patients undergoing antiviral treatment, were much higher than previously estimated (Table 2). Zidovudine treatment has previously been shown to decrease the amount of cell-free circulating HIV virus initially both in plasma of infected humans [19] and of severe combined immunodeficiency infected mice [22]. The high concentration of cell-free circulating HIV in our patients, who have been undergoing zidovudine therapy for over 10 months, may imply that some resistant viral strains have emerged.

The most striking feature of the results is the high level of cell-free HIV found in some CDC group II patients (patient 56, 70 and 82; Table 2). Serial samples from one CDC group II patient (patient 82) without antiviral treatment showed persistently high levels of plasma virus for several years and rapid turnover of sequence variants (Simmonds *et al.* in preparation). The detection of high levels of cell-free HIV from the plasma of both CDC group II and IV patients suggests that viral replication occurs continuously throughout the course of an HIV infection. We find no evidence for a virological 'latent' period.

Detection of HIV RNA from factor VIII

Two out of eight batches of factor VIII concentrate were positive for HIV-1 RNA by PCR; in one case with *env* primers and the other with *pol* primers. Both were confirmed by sequencing the PCR product. The amount of HIV RNA present in factor VIII is very low (2.5 copies per ml), and close to the threshold of detection. These two factor VIII sequences, which were both found in commercial products, are distinct from those of any published HIV isolates, but are more closely related to North American strains than to African ones.

We believe that this is the first direct demonstration of contamination of factor VIII by HIV-1. Previous studies by ourselves [23] and others [24] have detected hepatitis C virus (HCV) RNA sequences in factor VIII using similar methods. In fact, out of the eight batches of factor VIII concentrate tested for HIV-1 RNA in this study, all six of the commercially-derived batches also contained detectable amounts of HCV RNA, including the two positive for HIV RNA (the two National Health

Service-derived batches were negative for HCV RNA). The amounts of HCV RNA were higher than HIV; in two batches that were quantified by limiting dilution, between 30 000 and 100 000 copies of HCV RNA per ml were found [23]. There are many possible explanations for the difference in the concentration of HIV and HCV RNA. Comparison of the levels of HIV viraemia with that of HCV shows that HCV is present in 10–100-fold greater amounts in infected individuals. Second, the prevalence of HCV infection may be higher in paid donors. Third, HIV may be less stable during the factor VIII fractionation process than HCV, or may be excluded with greater efficiency. The last possibility could be the most likely if most plasma HIV is held in immune complexes.

We note that the *pol* sequence obtained from batch no. 8 contains an arginine residue at position 70 in the RT domain. This substitution was one of four found to be associated with resistance to zidovudine [25, 26]. The discovery of this substitution in factor VIII concentrate prepared before the use of zidovudine as an antiviral agent suggests that it was a pre-existing polymorphism.

Acknowledgements

We are grateful to Dr G. Kemball-Cook (National Institute for Biological Standards and Control) for supplying samples of factor VIII batches. We thank S. Rebus and D. Beatson for p24 antigen and CD4+ lymphocyte counts in the haemophiliac samples. These samples were collected by staff at the Haemophilia Centre, Royal Infirmary. We also thank Dr R. Al-Shawi for supplying cloned *tk* sequences and oligonucleotide primers, Dr P. Balfe for criticism of the manuscript and assistance in the early phases of the work, Dr S. Burgess for assistance with virus culture, materials and assays, and Mrs F. McOmish and Mr A. Cleland for technical assistance.

References

1. OU C-Y, KWOK S, MITCHELL SW, *ET AL.*: DNA amplification for direct detection of HIV-1 in DNA of peripheral blood mononuclear cells. *Science* 1988, 239:295–297.
2. KWOK S, MACK DH, MULLIS KB, *ET AL.*: Identification of human immunodeficiency virus sequences by using *in vitro* amplification and oligomer cleavage detection. *J Virol* 1987, 61:1690–1694.
3. SAIKI RK, GELFAND DH, STOFFEL S, *ET AL.*: Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 1988, 239:487–491.
4. WILLIAMS P, SIMMONDS P, YAP PL, *ET AL.*: The polymerase chain reaction in the diagnosis of vertically transmitted HIV infection. *AIDS* 1990, 4:393–398.
5. SIMMONDS P, BALFE P, PEUTHERER JF, *ET AL.*: Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy number. *J Virol* 1990, 64:864–872.
6. OKA S, URAYAMA K, HIRABAYASHI Y, *ET AL.*: Quantitative analysis of human immunodeficiency virus type 1 DNA in asymptomatic carriers using the polymerase chain reaction. *Biochem Biophy Res Commun* 1990, 167:1–8.
7. HART C, SPIRA T, MOORE J, *ET AL.*: Direct detection of HIV RNA expression in seropositive subjects. *Lancet* 1988, ii:596–599.
8. BYRNE BC, LI JJ, SNINKSY J, POIESZ JB: Detection of HIV-1 RNA sequences by *in vitro* DNA amplification. *Nucl Acids Res* 1988, 16:4165.
9. ARRIGO SJ, WEITSMAN S, ROSENBLATT JD, CHEN ISY: Analysis of *rev* gene function on human immunodeficiency virus type 1 replication in lymphoid cells by using a quantitative polymerase chain reaction method. *J Virol* 1989, 63:4875–4881.
10. LUDLAM CA, TUCKER J, STEEL CM, *ET AL.*: Human T-lymphotropic virus type III (HTLV-III) infection in seronegative haemophiliacs after transfusion of factor VIII. *Lancet* 1985, ii:233–236.
11. SIMMONDS P, LAISON FAL, CUTHBERT R, *ET AL.*: HIV antigen and antibody detection: variable responses to infection in the Edinburgh haemophiliac cohort. *BMJ* 1988, 296:593–598.
12. CHOMCQYNSKEI P, SACCHI N: Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 1987, 162:156–159.
13. CHURCH GM, GILBERT W: Genomic sequencing. *Proc Natl Acad Sci USA* 1984, 81:1991–1995.
14. WINSHIP PR: An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. *Nucl Acid Res* 1989, 17:1266.
15. BALFE P, SIMMONDS P, LUDLAM CA, BISHOP JO, LEIGH BROWN AJ: Concurrent evolution of human immunodeficiency virus type 1 in patients infected from the same source: rate of sequence change and low frequency of inactivating mutations. *J Virol* 1990, 64:6221–6233.
16. SIMMONDS P, BALFE P, LUDLAM CA, BISHOP JO, LEIGH BROWN AJ: Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. *J Virol* 1990, 64:5840–5850.
17. AL-SHAWI A, BURKE J, JONES CT, SIMONS JP, BISHOP JO: A Mup promoter-thymidine kinase reporter gene shows relaxed tissue-specific expression and confers male sterility upon transgenic mice. *Mol Cell Biol* 1988, 8:4821–4828.
18. ZACK JA, ARRIGO AJ, WEITSMAN SR, *ET AL.*: HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile latent viral structure. *Cell* 1990, 61:213–222.
19. HO DD, MOUDGIL T, ALAM M: Quantitation of human immunodeficiency virus type 1 in the blood of infected persons. *N Engl J Med* 1989, 321:1621–1625.
20. COOMBS RW, COLLIER AC, ALLAIN JP, *ET AL.*: Plasma viremia in human immunodeficiency virus infection. *N Engl J Med* 1989, 321:1626–1631.
21. MORROW WJM, WHARTON M, STRICKER RB, LEVY JA: Circulating immune complexes in patients with acquired immune deficiency syndrome contain the AIDS-associated retrovirus. *Clin Immunol Immunopathol* 1986, 40:515–524.
22. MCCUNE JM, NAAMIKAWA R, SHIH C-C, RABIN L, KANESHIMA J: Suppression of HIV infection in AZT-treated SCID-hu mice. *Science* 1990, 247:564–566.
23. SIMMONDS P, ZHANG LQ, WATSON HG, *ET AL.*: Detection, quantitation and sequencing of hepatitis C virus in haemophiliacs, drug abusers and blood products. *Lancet*, 1990 335:1469–1472.
24. GARSON JA, PRESTON FE, MAKRIS M, *ET AL.*: Detection of PCR of hepatitis C virus in factor VIII concentrates. *Lancet* 1990, 335:1473
25. LARDER BA, KEMP SD: Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science* 1989, 246:1155–1158.
26. LARDER BA, KELLAM P, KEMP SD: Zidovudine resistance predicted by direct detection of mutations in DNA from HIV-infected lymphocytes. *AIDS* 1991, 5:137–144.

THE CURRENT OPINION SERIES

CLINICAL MEDICINE

Current Opinion in:	Personal (circle one)	Institutional (circle one)
☐ Anaesthesiology (Vol 4, 6 issues)	£75/$145	£150/$255
☐ Cardiology (Vol 6, 6 issues)	£60/$100	£100/$170
☐ Dentistry (Vol 1, 6 issues)	£60/$100	£100/$170
☐ Gastroenterology (Vol 7, 6 issues)	£60/$100	£100/$170
☐ Infectious Diseases (Vol 4, 6 issues)	£60/$100	£100/$170
☐ Lipidology (Vol 2, 6 issues)	£65/$125	£115/$195
☐ Neurology & Neurosurgery (Vol 4, 6 issues)	£85/$145	£150/$255
☐ Obstetrics & Gynecology (Vol 3, 6 issues)	£70/$125	£125/$215
☐ Oncology (Vol 3, 6 issues)	£85/$145	£160/$270
☐ Ophthalmology (Vol 2, 6 issues)	£120/$195	£200/$340
☐ Orthopaedics (Vol 2, 6 issues)	£90/$145	£180/$290
☐ Pediatrics (Vol 3, 6 issues)	£60/$100	£100/$170
☐ Psychiatry (Vol 4, 6 issues)	£85/$145	£130/$230
☐ Radiology (Vol 3, 6 issues)	£85/$145	£170/$290
☐ Rheumatology (Vol 3, 6 issues)	£60/$100	£100/$170

BIOLOGICAL SCIENCES

☐ Biotechnology (Vol 2, 6 issues)	£200/$350	£300/$500
☐ Cell Biology (Vol 3, 6 issues)	£65/$120	£150/$275
☐ Genetics & Development (Vol 1, 4 issues)	£45/$80	£100/£185
☐ Immunology (Vol 3, 6 issues)	£65/$120	£150/$275
☐ Neurobiology (Vol 1, 4 issues)	£45/$80	£100/$185
☐ Structural Biology (Vol 1, 6 issues)	£90/$175	£190/$350

Method of Payment

SB

Total Order Value: .....

☐ Cheque enclosed, payable to Current Science Ltd

☐ Visa ☐ MasterCard ☐ Am Ex

Card No [ ] Exp Date [ ]

Name:

Address:

Postcode: Country:

Signature ..... Date .....

No risk guarantee

Your money will be refunded in full if you write cancelling your journal subscription within 30 days of receiving the first issue.

SEND TO:

USA & Canada

Current Science
20 North 3rd Street,
Philadelphia, PA 19106-2113,
USA Tel: (215) 574 2266

Rest of world

Current Science Ltd
34-42 Cleveland Street,
London, W1P 5FB UK
Tel: (071) 323 0323

US$ prices applicable to subscribers in USA & Canada only

Subscription rates include airspeed delivery.

Subscriptions run on a calendar year basis.

☐ Please send me an institutional invoice

☐ Please send me a sample copy of the title(s) circled above

☐ 30% discount for Students/ Housemen/ Residents. Deduct 30% from the prices above and provide proof of status

AIEX

☐ Bank transfer (details available upon request)

☐ Please send me information on your electronic databases - available for all titles

JOURNAL OF VIROLOGY, Nov. 1991, p. 6266–6276
0022-538X/91/116266-11$02.00/0

Vol. 65, No. 11

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

PETER SIMMONDS,[1]* LIN QI ZHANG,[2] FIONA McOMISH,[2] PETER BALFE,[2] CHRISTOPHER A. LUDLAM,[3] AND ANDREW J. LEIGH BROWN[2]

Department of Medical Microbiology, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG,[1] Institute of Cell, Animal and Population Biology, Division of Biological Sciences, University of Edinburgh, Kings Buildings, Edinburgh, EH9 3JN,[2] and Department of Haematology, Royal Infirmary of Edinburgh, Lauriston Place, Edinburgh, EH3 9YW,[3] Scotland

Received 10 June 1991/Accepted 13 August 1991

Sequence change in different hypervariable regions of the external membrane glycoprotein (gp120) of human immunodeficiency virus type 1 (HIV-1) was studied. Viral RNA associated with cell-free virus particles circulating in plasma and proviral DNA present in HIV-infected peripheral blood mononuclear cells (PBMCs) were extracted from blood samples of two currently asymptomatic hemophiliac patients over a 5-year period. HIV sequences were amplified by polymerase chain reaction to allow analysis in the V3, V4, and V5 hypervariable regions of gp120. Rapid sequence change, consisting of regular replacements by a succession of distinct viral populations, was found in both plasma virus and PBMC provirus populations. Significant differences between the frequencies of sequence variants in DNA and RNA populations within the same sample were observed, indicating that at any one time point, the predominant plasma virus variants were antigenically distinct from viruses encoded by HIV DNA sequences in PBMCs. How these findings contribute to current models of HIV pathogenesis is discussed.

The high degree of sequence variability that exists between different isolates of human immunodeficiency virus type 1 (HIV-1) (1, 38) poses a major problem for the development of effective methods of immunization against the virus. In particular, a major site for antibody-mediated virus neutralization in the *env* gene (the V3 hypervariable region [15, 17, 22, 25, 27]) shows considerable sequence heterogeneity and rapid rates of sequence change (1, 3, 12, 20, 36, 38, 43, 46). Furthermore, many of the amino acid changes in this region have been shown to modulate immunological recognition (22, 24, 30, 40).

We have used phylogenetic analysis of nucleotide sequences from a set of five serial samples from a (currently) asymptomatic hemophiliac patient infected with HIV-1 to investigate the rate and direction of sequence change in each of three hypervariable regions (V3, V4, and V5 [27]). By using a nested polymerase chain reaction (PCR) to amplify viral nucleic acids in vivo (37, 49), sequences of proviral DNA from peripheral blood mononuclear cells (PBMCs) were compared with those of viral RNA in plasma. We observed significant differences between the two populations in all three hypervariable regions at different points after infection. We present and discuss a model of HIV pathogenesis that takes these results into account along with the results of previous investigations of biological heterogeneity of HIV (2, 6, 41), the cell types infected with HIV in vivo (34, 35), and the evidence for positive selection for sequence change in hypervariable regions of the *env* gene (5, 36, 46).

* Corresponding author.

MATERIALS AND METHODS

Patient samples. Sequential samples from a hemophiliac patient, p82, infected with HIV-1 from factor VIII prepared from Scottish blood donations in 1983 (23), were used for sequential studies of HIV sequence change. Seroconversion took place in June 1984, at which time a plasma sample was stored. Subsequent samples (from both plasma and PBMCs) from this patient were collected in June 1987, January 1988, February 1989 (1989A), and August 1989 (1989B). Several of the batches of factor VIII transfused to p82 were given to another hemophiliac patient, p80, who also seroconverted in 1984. A PBMC sample from this second patient was taken in February 1989 and was used for sequence comparisons.

PCR product length analysis. Sequence variants that differed in length in the V4 and V5 hypervariable regions were visualized by high-resolution gel electrophoresis of amplified DNA (36, 45). For V4 sequences, proviral DNA or cDNA was amplified first with primers e and h and then, in a second PCR, with primer f and a new antisense primer lying in the C3 region (5′ ATGGGAGGGCATACATTGC; position 7539 in pHIVHXB2). To amplify V5 sequences, the second PCR was carried out with primer g and a new sense primer in the C3 region (5′ GGAAAAGCAATGTATGCCC; position 7515 in pHIVHXB2). The relative proportions of sequence variants of different molecular weights within a sample was obtained by replicate amplification of undiluted proviral DNA (or viral cDNA) samples containing typically 100 to 200 molecules of target sequence. Bands were quantified by scanning densitometry of the autoradiographic image of the polyacrylamide gel with a Shimadzu densitometer.

Sequencing of HIV gp120. Proviral DNA was extracted from PBMCs as previously described (37). Single molecules of provirus were amplified after prior limiting dilution and

directly sequenced to avoid errors introduced by the amplification of DNA in vitro (3, 36). Viral RNA was extracted from plasma, reverse transcribed, and sequenced as previously described (49). The region amplified spanned the V3, V4, and V5 hypervariable regions (27) and was amplified with primers a to h in three sequential nested PCRs as previously described (36). This method gave continuous sequence from nucleotides equivalent to positions 6957 to 7814 in clone pHXB2 of $HIV_{HTLV\text{-}IIIB}$. Sequences from V3, V4, and V5 and surrounding nucleotides are presented in this article. A fuller analysis of the V3 sequences is in preparation.

Evolutionary analysis of gp120 sequences. Sequences obtained in this and our previous study were aligned by the Needleman and Wunsch algorithm as implemented by the program GAP on the University of Wisconsin GCG Package (8) and subsequently edited by hand. A matrix of evolutionary distances was generated by using the two-parameter model of Kimura (18); alignments that gave the minimum evolutionary distance between sequences were used in this study. Phylogenetic trees were constructed on the basis of the distance matrices by using the Fitch-Margoliash procedure (10) as available in program FITCH of the PHYLIP package supplied by J. Felsenstein. The validity of the trees was assessed by reentering the tree obtained by FITCH into the maximum-likelihood-based program DNAML as a user-defined tree. This gave confidence intervals for each of the internodal distances.

Nucleotide sequence accession numbers. Sequences obtained in this study have been submitted to GenBank under accession numbers M77541 through M77636.

RESULTS

Nucleotide sequence variation in V4 and V5. A longitudinal study of HIV sequence change was carried out with samples from p82, a hemophiliac patient infected with HIV-1 from factor VIII concentrate. Sequences in the V4 and V5 regions were obtained from stored plasma from this individual at the time of acute seroconversion (May 1984) and subsequently from plasma and PBMCs at each of four time points following infection (June 1987, January 1988, February 1989 [1989A], and August 1989 [1989B]). This individual remained asymptomatic during the course of the study and has not received zidovudine or other antiviral therapy at any time. The results of standard virological and immunological investigations of this individual are shown in Table 1.

For comparative purposes, proviral DNA in a PBMC sample from another hemophiliac patient, p80, taken 5 years after infection, was also sequenced. This patient was infected by contaminated factor VIII around the same time as p82 (May 1984) and has also remained asymptomatic over the study period.

A total of 86 sequences in the V4 hypervariable region and 70 in the V5 region of *env* were obtained from p82; 9 V4 and 9 V5 sequences were obtained from the single PBMC sample from p80. The sequences obtained over the 5-year course of the longitudinal study were highly variable in the V4 and V5 regions, indicating rapid and continuous sequence change over the asymptomatic period of the infection. Each V4 and V5 nucleotide sequence was aligned against all the others by using a standard algorithm (GAP; see Materials and Methods), and a nucleotide distance matrix was obtained from pairwise comparisons. The evolutionary relationships between the different sequences were estimated by the FITCH program. In the resulting phenograms (Fig. 1a and b), evolutionary distances are shown by the horizontal separation between pairs of sequences (the vertical lines are of no significance).

TABLE 1. Standard virological and immunological markers of HIV infection in p82 and p80

Sample	Time (months)[a]	$CD4^+$ lymphocytes (10^9)	p24 antigen[b]	Provirus-bearing PBMCs[c]
p82				
1983	−14	1.45	NA	NA
1984	0	0.93	+	ND
1987	36	0.53	−	1/2,000
1988	43	0.34	−	1/2,270
1989A	56	0.65	+	1/700
1989B	63	0.16	−	1/800
p80 (1989)	32	0.42	−	1/50,000

[a] Time from first positive serum sample.
[b] Detection of serum antigen by capture enzyme-linked immunosorbent assay (ELISA) (>15 pg/ml; Dupont). NA, not applicable.
[c] Proportion of PBMCs bearing provirus, estimated by limiting dilution (37). ND, not done.

A notable feature of this analysis is the apparent clustering of sequence variants into a small number of groups. In the V4 region (Fig. 1a), three groups can readily be identified (A, B, and C). Only one sequence, lying between groups A and C, does not fit into the classification. Clusters of distinct sequence types are also discernible in the V5 region (Fig. 1b), although in this case there are more groups (here labelled A to E) and some sequences that do not fit any of the groups (indicated by ?). Sequence variation within the V4 and V5 groups is considerably less than that which exists between groups. Unexpectedly, some of the sequences from p80 were identical to those of p82 in the V4 region (group A), while some had diverged to form a group clearly distinct from V4A, -B, or -C (Fig. 1a). Similarly, some of the V5 sequences from p80 were identical to those of p82 (V5A), while others fit none of the other p82 groups (Fig. 1b). As p80 and p82 shared several batches of noncommercial factor VIII in the year prior to seroconversion, and in view of the presence of identical V4 and V5 sequences in both, we infer that they were infected from the same source. Sequences of the V4A and V5A type are likely to have formed a major component of the virus population that infected both patients.

The mean within-group distances were 4.6, 5.6, and 1.0% for the V4A, -B, and -C groups, respectively, while the mean intergroup distances ranged from 11.0% (V4A to V4C) to 30.4% (V4B to V4C). In the V5 region, intergroup distances ranged from 16 to 55%, while within-group variability in no case exceeded 6.5%. In the V4 region, the branching pattern suggests that V4B and V4C sequence types diverged independently from V4A, the group that contains sequences found at seroconversion and those that are shared between p80 and p82. The major V5 groups also appear to have evolved independently from V5A. However, the distances between groups are so large that definite conclusions about such relationships cannot be made with these sequence data.

Amino acid sequence variation in V4 and V5. Figure 2 shows the translated sequences in the V4 region, divided into the groups indicated by the phenogram, to illustrate the differences between sequences within groups and the much greater differences between groups. The consensus sequences of V4A, -B, and -C are clearly distinct from each

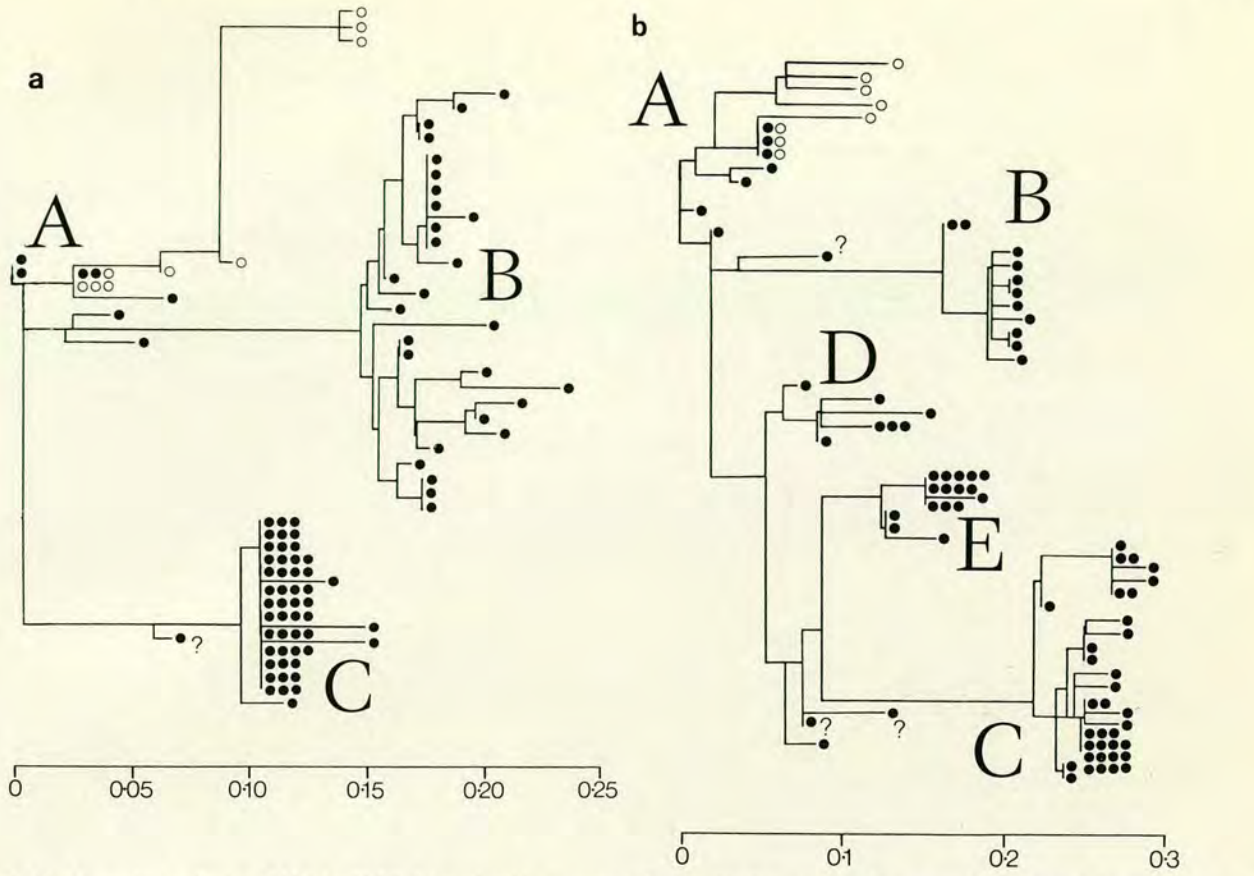


FIG. 1. Phylogenetic analysis of V4 (a) and V5 (b) nucleotide sequences from p80 (○) and p82 (●) over a 5-year period from the time of seroconversion. Sequence types are indicated as A to C (V4) or A to E (V5); intermediate and unclassified sequences are indicated by ?. Evolutionary distances between pairs of sequences are proportional to horizontal separation, as indicated on the scale. Maximum-likelihood analysis indicated that the distances between all nodes represented were significantly different from 0.

other, whereas in each of the three groups, individual sequences rarely differ from each other by more than two to three amino acid residues. Similarly, the potential sites for N-linked glycosylation in the hypervariable regions differ considerably between groups (indicated by # in Fig. 2). Within groups, the number and spacing of sites are relatively constant, although V4B is more variable in this respect, containing similar numbers of variants that differ at one or two of the four potential sites in the region.

The V4 region is known to show considerable variability in length among different isolates (1, 12, 27, 38) and also between proviral sequence variants within a single sample (36, 45). Comparison of the consensus sequences in this region from p82 (Fig. 2) reveals that virtually all V4B variants consist of 17 amino acid residues between the relatively more-conserved flanking regions (..FNSTW----<V4>----ITLPCR...), with only three exceptions (sequences 7, 8, and 16) which are two or three amino acids longer. Similarly, the lengths of all but three (2, 3, and 4) of the V4C sequences are constant, at 18 amino acids between the conserved flanking regions.

The lengths of the V4A sequences are somewhat more variable. Sequences of type 1 of V4A from p82 and the four sequences of type 1 of p80 are identical at both the amino acid and nucleotide levels and are all 15 amino acids long, the same length as V4A-2 and V4A-3. However, in both individuals, longer V4 sequences are also found: in p82, there were three additional sequences of 20, 22, and 24 residues (sequences 5, 4, and 3 respectively), and in p80, there were four sequences of 27 amino acids and one of 26.

Variation in the V5 region (Fig. 3) shows many of the characteristics of variation seen in V4. The consensus sequences of the five groups (V5A to E) differ considerably from one another, while sequence variation within groups is minimal, particularly in groups B to E. Group A sequences, defined by the phylogenetic tree (see above), appear to contain two types of sequences at the amino acid level (sequences 4 and 5 appear distinct from the others), although for the purposes of analysis (see below), the numbers of sequences are so small as not to justify further subgrouping. The pattern of N-linked glycosylation sites is also well conserved within groups, and the overall lengths of the regions (between ...TRDGG---<V5>---FRPGG....) are 7 to 10, 8, 12, 8, and 8 residues in groups A to E, respectively. Sequence type 1 from p80 (n = 3) in the V5 region is identical to V5A-1 (n = 3) of p82 (Fig. 3), while the other relatively small number of other sequences from p80 differ considerably from the common type and from each other. Further analysis would be necessary to find out whether sequences from this patient grouped into distinct types as they appear to do in this region from p82.

Sequence change in the V4 and V5 regions. Having defined and analyzed the sequence groups in the two hypervariable regions, the classification can be used to study sequence

p80)

p80V4-1	. # -# ----------# k	4
-2	. # -# iq # #	1
-3	. # s# iq # # d	1
-4	. # f# # # #d	2
-5	. # f# # # #	1
Con	LFNSTWNSTQL-NSTW**tstllnstwnn**NST**eet**ITLPCR	9

A)

V4A-1	..... # # # --------- k	2
-2	..... # # # --------- v kt	1
-3	# # # #n--------- #	1
-4	. # # # tqlnsag#n #	1
-5	# # f# # tqlnsar-- #l	1
-6	. # # i f# # tqlnsa---- #	1
Con	LFNSTWNS**t**Q**l**NSTWN**s**---------**te**E**ni**TLPCR	7

B)

V4B-1	.... #--- #-- s #	6
-2	..# #--- #-- s s #	1
-3	..... #--- #-- s g #	1
-4	# #--- #-- t #	1
-5	# s--- #-- s g #	1
-6	. # s--- #-- s k g #	1
-7	# w#- #-- s g #	1
-8	... w#y #d -- s #	1
-9	. # #--- # -- s #	1
-10	..... #--- # -- i #	1
-11	# #--- # -- # g #	2
-12	# s--- # -- # g #	1
-13	# #--- #d -- n i #	1
-14	. # t--- #d -- # g #	1
-15	. # t--- #d -- # r g #	1
-16	. # t--- #d yt # g #	1
-17	. # v--- # -- # k #	1
-18	. # i--- # -- # g #	1
-19	# --- y# -- # g #	4
Con	LFNSTW**n**---Y**sngt**--W?S**t**Q**hn**T**e**ENITLPCR	28

C)

V4C-1	# # # # #	35
-2	. # # # -- #	1
-3	# # ---# #	1
-4	. # # ---- #	1
-5	# # p# #	2
-6	. # # #n #	1
-7	# # #g #	1
-8	. # # # d #	3
-9	# # # d kd#	2
-10	. # # f# #	1
-11	.... # f# p #	1
-12	. # # i r #	1
Con	LFNSTWNSTWDL**tqlnstqnkee**NITLPCR	50

UNCLASSIFIED:

V4o-1	....TW#STQP#STRNNNEE#ITLPCR	1

FIG. 2. Peptide sequences of the three phylogenetic groups in the V4 hypervariable region. Con, consensus sequence for each group; nonconserved residues are shown in boldface lowercase letters. Differences from consensus are shown for each sequence entry, and frequency of detection of each sequence type is shown in the rightmost column. Symbols: ?, no majority at this position;., unsequenced; –, gap introduced to preserved sequence alignment within group; #, all potential sites of N-linked glycosylation (nonconserved sites shown in boldface).

p80)

p80V5-1	# ----s t i	3
-2	enkp-d t t .	1
-3	e#ttk# t t	1
-4	i #---ktt t	1
-5	trqdr-d t t	1
-6	i -er-dp- il ...	1
Con	GLLLTRDGG**ng**----?**et**E?**f**RPGGG	9

A)

V5A-1	#-- -	3
-2	#--k -	1
-3	#--k d-	1
-4	r#e #t	1
-5	k#e #t	1
Con	GLLLTRDGGN--g**S**e-TEIFRPGGG	7

B)

V5B-1	# .	3
-2	s# .	1
-3	# t .	1
-4	# #g	5
-5	- # t .	1
Con	GLLLTRDG**g**N**n**T**ert**E**i**FRPGGG	11

C)

V5C-1	# #	17
-2	d #	1
-3	# # #	1
-4	# # i	2
-5	#--- ..	1
-6	#--- k ..	1
-7	i g # #	1
-8	g # #	3
-9	g # s# i	7
-10	g qrdm s i .	1
Con	GLL1TRDGGN**s**G**nksnd**T**t**E**t**F**r**PGGG	35

D)

V5D-1	####	3
-2	k##	1
-3	k## t k ..	1
-4	q##	1
-5	qd# t	1
Con	GLLLTRDGGN**rn**NTTE**i**F**r**PGGG	7

E)

V5E-1	#	11
-2	# k ..	2
-3	# i	2
-4	# ip	1
Con	GLLLTRDGGDTSNTTE**ifr**PGGG	16

UNCLASSIFIED:

V5o-1	LLLTRDGG#KSKNDPETPRPGGG	1
V5o-2	LLLTRDGG#TSTTEIFRPG..	1
V5o-3	LLLTRDGGNR##TTETFRPGGG	1
V5o-4	LLLTRDGG#TSKTTEIFRPGGG	1

FIG. 3. Peptide sequences of the five phylogenetic groups in the V5 hypervariable region. Arrangement and symbols are same as for Fig. 2.

changes over time in samples from p82. Figure 4a records the numbers of sequences detected in plasma (above the *x* axis), and PBMCs (below the *x* axis). Although the numbers of sequences obtained at any one time point are relatively low, there is clear evidence for turnover of sequence variants. Type A variants were found in all four of the seroconversion RNA sequences in 1984, while only sequences of type B were found in 1987. In the following year, the most commonly found sequence was type C, which appears to have completely replaced type B in the two samples taken in 1989. Turnover of sequence variants may also be observed in the provirus population. Both V4A and V4B sequences were found in 1987, while in the following year almost all sequences were of type B. The replacement of V4B with V4C was completed in the following year.

Comparable turnover of sequence variants is also found in

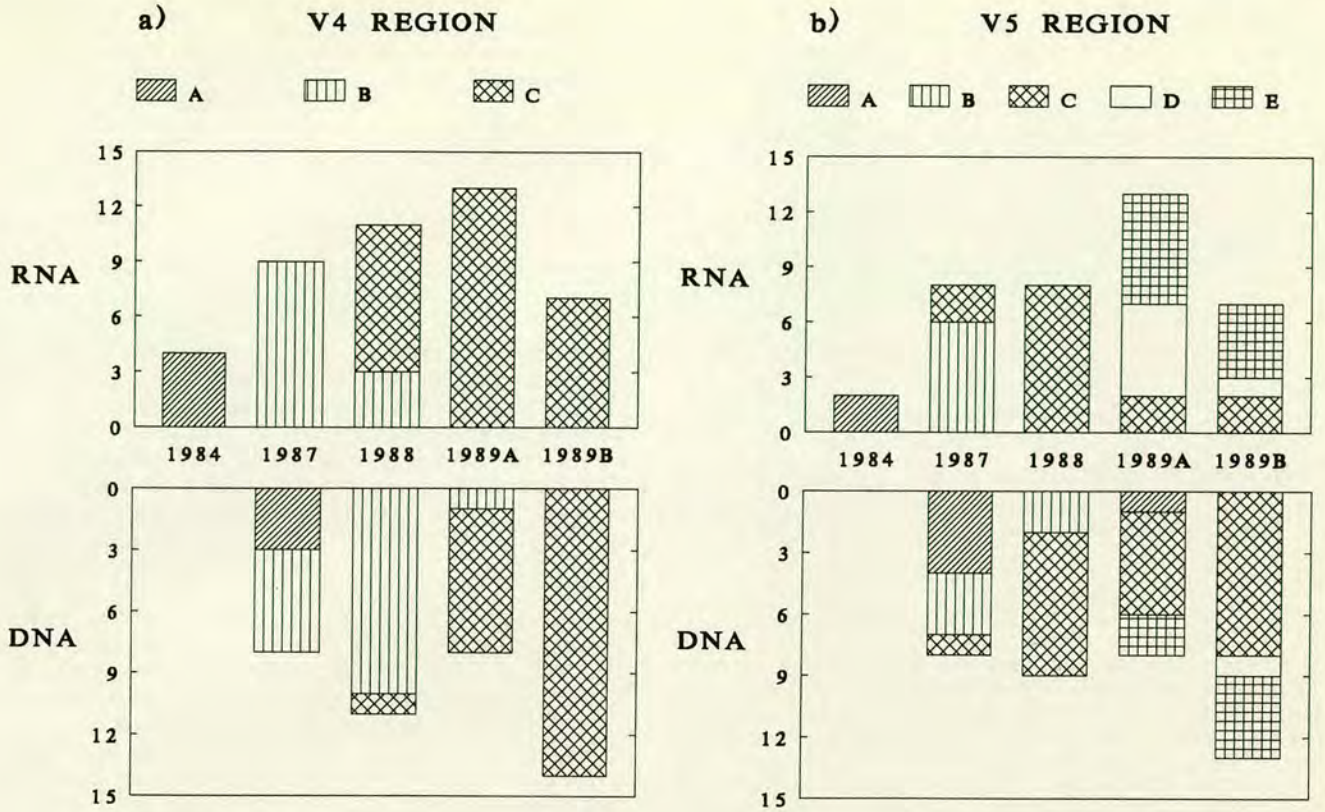


FIG. 4. Frequency of detection of V4 (a) and V5 (b) sequence types in sequential samples from p82 (1984 to 1989B). RNA sequences are shown above the *x* axis, and DNA sequences are shown below the *x* axis.

the V5 hypervariable region (Fig. 4b). Whereas V5A sequences were found at seroconversion in the plasma, these were replaced successively by V5B in 1987, V5C in 1988, and finally by V5D and -E in 1989. A similar progression was also observed in the proviral population. In 1987, approximately half of the sequences were of seroconversion type A. The almost complete replacement of these sequences by V5B and V5C took place in the following two years. In turn, V5D and -E appeared to be in the process of replacing V5C by the end of the study period.

Differences between DNA (proviral) and RNA (viral) populations. At several time points, there were considerable differences in the frequencies of different sequence types in the DNA and RNA samples. For example, the 1988 DNA sample contained predominantly V4B sequences in PBMCs (10 of 11) yet mainly V4C sequences in the plasma samples (8 of 11). Similarly, the preponderance of V5C sequences in the two 1989 DNA samples (6 of 9 and 8 of 13) contrasted with the infrequency of their detection in the corresponding plasma samples (2 of 13 and 2 of 7). The appropriate statistical procedure for comparing frequencies in small samples, Fisher's exact test, was used to test the significance of these differences. The frequencies of V4 variants in the 1988 sample and of V5 variants in the 1989A sample were found to be significantly different between the PBMC proviral and plasma RNA populations ($P < 0.01$ and $P < 0.05$, respectively).

The relative frequencies of the various sequence types in the V4 and V5 regions was also estimated by high-resolution gel electrophoresis of amplified DNA (36, 45). As indicated previously (Fig. 2), 25 of the 28 V4B sequences had an overall length of 17 amino acids, while 47 of the 50 V4C sequences were 18 amino acids long. Aliquots of DNA (2 μg) extracted from the PBMC samples in 1988 and 1989A containing approximately 70 and 220 molecules of provirus (Table 1) and undiluted cDNA after reverse transcription of RNA sequences present in plasma (containing 100 to 200 copies of target sequence; data not shown) were amplified in two stages with primers specific for the V4 region (see Materials and Methods). The product DNA consisted of two size variants, differing in electrophoretic mobility by 3 bp (Fig. 5). As indicated, the smaller band corresponds to the predicted size of V4B and the larger band corresponds to V4C. The 1988 DNA sample (lanes b) consists of mainly V4B sequences, while the corresponding RNA sample (lanes c) consists of predominantly V4C. The almost complete replacement of V4B sequences by V4C between 1988 and 1989 (Fig. 4) is also shown by this length analysis; both DNA (lanes d) and RNA (lanes e) contain predominantly V4C sequence types.

The relative numbers of V4B and V4C sequences were quantified by scanning densitometry (Table 2). To show that representative numbers of sequence variants had been amplified, each sample was amplified in replicate to allow two independent samplings of the populations present. The reasonably close agreement between all of the replicates confirmed that the populations studied (>100 sequences in each sample) were sufficiently large to prove that the differences between the populations at both time points were not due to sampling error. Furthermore, the relative proportions agree closely with those determined by sequence analysis (Fig. 4). For example, the 1988 DNA sample contained 74 to 75%

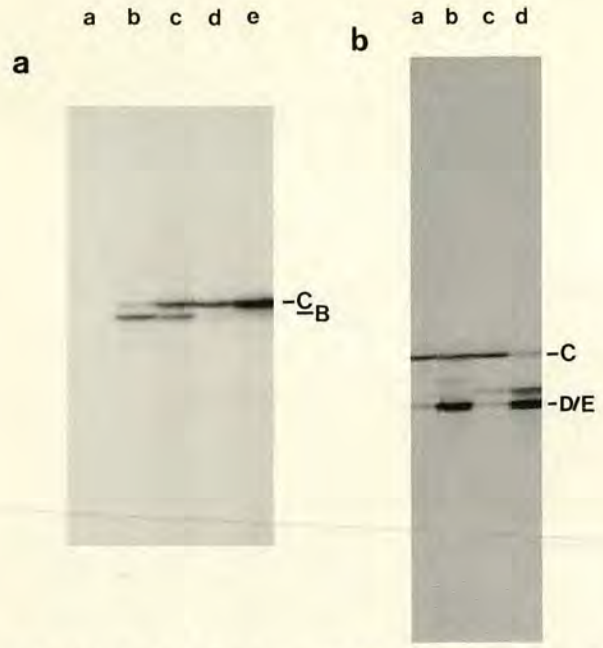


FIG. 5. Length analysis of amplified DNA in the V4 (a) and V5 (b) hypervariable regions to confirm existence of population differences in the in vivo DNA and RNA populations. (a) Lanes: a, negative human DNA amplified with primers spanning the V4 region; b and d, PCR product from V4 region of proviral DNA from the 1988 and 1989A samples, respectively, from p82; c and e, PCR product from viral RNA in the corresponding plasma samples. Expected sizes of V4B and V4C sequences are indicated. (b) Lanes: a and c, PCR product from V5 region of proviral DNA from the 1989A and 1989B samples, respectively; b and d, corresponding RNA samples. Expected sizes of V4C and V4D and -E are indicated.

V4B sequences while the RNA sample contained only 42 to 48%. The corresponding numbers of V4B sequences are 10 of 11 and 3 of 11 in these two samples. Similarly, the 1989 DNA sample contained 84 to 85% V4C sequences by densitometry, compared with 10 of 11 by sequence analysis, while the RNA population was uniformly V4C by both methods.

TABLE 2. Serial changes in the frequencies of V4 and V5 length variants in plasma (RNA) and PBMC (DNA) samples from p82 estimated by densitometry

Sample	Year[a]	Variant[b]					
		V4B (%)	V4C (%)	Ratio	V5C (%)	V5D, -E (%)	Ratio
DNA	1988	75, 74	25, 26	2.92	ND[c]	ND	ND
RNA	1988	48, 42	52, 58	0.82	ND	ND	ND
DNA	1989A	15, 16	85, 84	0.18	77, 63	23, 37	2.33
RNA	1989A	0, 0	100, 100	0.00	32	68	0.47
DNA	1989B	ND	ND	ND	75, 78	25, 22	3.26
RNA	1989B	ND	ND	ND	18	82	0.22

[a] 1989A, samples were taken in February 1989; 1989B, samples were taken in August 1989.

[b] Percents are replicate densities of independently amplified aliquots of the original DNA or cDNA. Repeat samples were not available from RNA samples in the V5 region.

[c] ND, not done.

TABLE 3. Frequencies of the combinations of V4 and V5 types in the 75 complete sequences obtained in the study

V5 sequence type	V4 sequence type		
	A	B	C
A	5	2	
B		11	
C		9	25
D			7
E			16

An equivalent analysis of the V5 region was carried out, and the discrepancy between the relative numbers of (i) V5C and (ii) V5D and V5E in the two populations was investigated. V5C differs in length from V5D and -E by 12 nucleotides (Fig. 3). Figure 5b confirms the existence of a marked difference in the relative numbers of the two sequence types in both the 1989A and 1989B samples. Ignoring the sequences that are of intermediate length between the two main types and whose classification is uncertain, the majority of DNA sequences are V5C in both the 1989A and 1989B samples (63 to 78%), which is comparable to the numbers of sequences found previously (6 of 9 and 8 of 13). By contrast, 68 to 82% of the corresponding RNA sequences were of type D or E, reflecting their frequency of detection by sequence analysis (11 of 13, 5 of 7) at the two time points.

Having established by two methods that significant differences exist between the two populations at at least three time points, a more detailed consideration of the origin of these differences is justified. A general trend that is found in both the V4 and V5 regions is for RNA sequences to turn over more rapidly than the corresponding DNA sequences (Fig. 3). For example, the seroconversion type V4A sequence is completely replaced in plasma by 1987 yet forms a substantial proportion of sequences in PBMCs at that time. Similarly, the difference in the relative numbers of V4B and -C sequences in the 1988 sample and the replacement of V5C with the V5D and -E variants in the 1989 sample could be interpreted as a more rapid transition to a new sequence type in the plasma. The possible mechanisms and the consequences of this observation are discussed below in relation to current models of HIV pathogenesis.

Linkage of V4 and V5 sequences. To obtain the sequences in this study, single molecules of HIV provirus or RNA reverse transcript were isolated by limiting dilution prior to amplification with primers spanning the entire V3-C2-V4-C3-V5 region. With this method, we have obtained sequences that are not only free of errors associated with copying of DNA in vitro but also have avoided the problems of producing sequences that are hybrids of two or more HIV sequences present in the patient sample because of switching between different templates during the amplification process (26). These sequences can therefore be used for studies of linkage and recombination in vivo.

A very restricted number of combinations of V4 and V5 sequences were observed in our datum set (Table 3). We found that there was no fixed relationship between a given V4 sequence type with those of V5. For example, HIV sequences of type V4A could contain either V5A or V5B sequences; similarly, V4B was associated with V5A, -B, and -C. Finally, as well as being linked to V5C, V4C was also found in viral sequences containing the V5D and V5E sequence types. A consequence of the variable associations

SAMPLE	n	10 20 30 36
1984 Plasma	1	N N T R K **S** I **H** I G P G **R A F** Y **T** T **G E** I **I G D** I R Q
Patient #80 (1989 PBMC)	8	N N T R K **S** I **H**$_4$ **P**$_2$ **N**$_2$ I G P G **R A F** Y **T**$_6$ **A**$_2$ T **G E**$_6$ **D**$_2$ I **I G**$_7$ **R**$_1$ **D** I R Q
1987 PBMC	5	N N T R K **S** I **H**$_4$ **P**$_1$ I G P G **R A F** Y **T** T **G E**$_3$ **Q**$_1$ **G**$_1$ I **I**$_3$ **M**$_1$ **V**$_1$ **G D** I R Q
1987 Plasma	9	N N T R K **S**$_8$ **G**$_1$ I **H**$_5$ **P**$_4$ I G P G **R**$_8$ **G**$_1$ **S**$_1$ **A F** Y **T**$_8$ **A**$_1$ T **G E**$_3$ **D**$_3$ **G**$_2$ I **I G D** I R Q
1988 PBMC	11	N N T R K **S**$_7$ **R**$_3$ **G**$_1$ I **H**$_8$ **P**$_3$ I G P G **R**$_9$ **S**$_1$ **G**$_1$ **A**$_{10}$ **T**$_1$ **F**$_8$ **V**$_3$ Y **T** T **G Q**$_5$ **G**$_4$ **D**$_2$ I **I G D**$_{10}$ **N**$_1$ I R Q
1988 Plasma	11	N N T R K **R**$_5$ **G**$_4$ **S**$_2$ I **H** I G P G **R**$_8$ **S**$_3$ **A V**$_7$ **F**$_4$ Y **T**$_8$ **A**$_3$ T **E**$_6$ **G**$_5$ **Q**$_7$ **G**$_3$ **R**$_1$ I **I G N**$_6$ **D**$_5$ I R Q
1989A PBMC	8	N N T R K **R**$_6$ **G**$_2$ I **H**$_4$ **Y**$_4$ I G P G **R**$_6$ **S**$_2$ **A V**$_6$ **F**$_2$ Y **T**$_6$ **A**$_2$ T **E**$_5$ **G**$_3$ **Q**$_6$ **G**$_2$ I **I G N**$_4$ **D**$_3$ I R Q
1989A Plasma	13	N N T R K **G**$_{11}$ **R**$_2$ I **H**$_{11}$ **Y**$_2$ I G P G **S**$_{12}$ **R**$_1$ **A F**$_{10}$ **V**$_3$ Y **A**$_{11}$ **T**$_2$ **A**$_1$ T **G**$_{11}$ **E**$_2$ **G**$_{10}$ **Q**$_2$ I **I G D**$_{11}$ **N**$_2$ I R Q
HIV-MN		Y N K R K R I H I G P G R A F Y T T K N I I G T I R Q

FIG. 6. Sequences at the center of the V3 loop in sequential samples from p82 and a single DNA sample from p80 (sequence of HIV_{MN} included for comparison). Numbering begins from the cysteine residue at the start of the V3 loop. Variable residues are shown in boldface type, numbers of the major and minor variants at variable sites are shown in subscript, and numbers of sequences obtained from each sample are indicated (n).

between hypervariable regions was that the frequencies of sequence types varied independently from each other. For example, it can be seen from Fig. 4 that the predominant virus type in the V4 region remained V4C at a time when V5 sequences were changing from V5C to V5D and -E. Similarly, while V4A sequences were being replaced by V4B in 1987 to 1988, V5 sequences underwent two replacements from V5A to V5B and then to V5C over the same interval. However, the changes in the frequencies precluded a statistical investigation of association between variants (linkage disequilibrium).

Combinations of V4 and V5 sequences showed a higher rate of turnover than that of the different sequences considered separately. This led to even greater differences between the DNA and RNA populations at a given time point. For example, in 1987, most of the RNA sequences were of combined type BB, while the DNA was predominantly BA. In 1988, RNA genotypes were almost exclusively CC, while those of DNA were mainly BC (data not shown). Significant differences between the frequencies of V4-V5 combinations between the DNA and RNA populations were found at all four time points (data not shown).

By using the data on the frequencies of V4 and V5 combinations, the following succession of genotypes was observed over the 5-year follow-up period:

$$AA \rightarrow BA \rightarrow BB \rightarrow BC \rightarrow CC \rightarrow CD$$
$$\searrow?$$
$$\rightarrow CE$$

This temporal succession of genotypes does not necessarily imply that each preceding form was ancestral to the variant that succeeded it. In the V4 region (Fig. 1), V4B and V4C appear as lineages entirely independent from V4A, although the accuracy of the phylogenetic analysis is limited by the extremely high rate of sequence change in this region. If the derived V4 and V5 sequences have evolved independently from the seroconversion type, then recombination must have occurred in vivo to generate the combinations of sequences found (Table 3).

Sequence change in the V3 region. The role of the V4 and V5 regions in antigenic recognition has not yet been defined. To investigate whether the difference in the sequence types in the DNA and RNA populations would lead to alterations in the susceptibility of the virus to antibody-mediated neutralization, a set of sequences similar to those of the V4 region were obtained in V3. In Fig. 6, we show the sequences obtained from these samples at the center of the V3 loop which include epitopes that have been implicated in both antibody-mediated and cytotoxic T-cell recognition (22, 40). The single sequence obtained from the plasma sample from p82 at seroconversion differed from that of HIV_{MN} at many of the sites shown. This sequence was identical to three of the five sequences in the 1987 DNA sample from p82 and to four of the eight sequences in the 1989 sample from p80. It is therefore likely to have formed a major component of the infectious virus population in the factor VIII given to both patients.

While the 1987 DNA and RNA samples were similar in V3 sequences, considerable differences between the DNA and

RNA samples in the frequencies of amino acids were observed in both the 1988 and 1989 samples from p82. For example, at residue 10 of the V3 loop, the majority of DNA sequences had an arginine in the 1989 sample, whereas the corresponding RNA sequences were generally glycine. Similar discrepancies were found at residues 17, 19, 21, 23, and 24. Substitutions at many of these residues have been previously shown to abolish serological or T-cell reactivity. Thus, most of the viruses encoded by the RNA sequences are probably quite different antigenically from those encoded by proviral sequences in the PBMCs. The significance of this finding for sequential studies of virus neutralization is discussed below.

DISCUSSION

Rate of sequence change in gp120. In this study, we have produced evidence for rapid sequence change in three hypervariable regions of *env*. This finding was anticipated by our own cross-sectional studies of sequence evolution in a cohort of hemophiliac patients infected from a common source (3, 36) and of the V3 sequences of six children infected from a single plasma donation (46). Although neither study determined the infecting sequence population, the existence of substantial sequence variation among individuals 3 to 5 years after infection allowed an estimate of the rate of sequence change from a calculated common ancestor (in terms of percent nucleotides per year) to be made (3). The model used for this calculation allows for differences in the rate of sequence change among individuals, and the nucleotide distance estimates are corrected for multiple substitutions (18), but this approach does not account for convergence of sequences due to selection. It also assumes a steady accumulation of substitutions with time.

We have shown here that sequence evolution in p82 was indeed more rapid than that in p80, but, more important, that substitutions do not accumulate steadily with time. Sequence change in p82 over the five years of follow-up consisted of a series of replacements of one particular sequence type with another. We have shown that succeeding sequence types were not necessarily directly derived from the previous sequence; for example, V4C succeeded V4B in 1988 to 1989, yet V4B may not be the immediate ancestor of any of the V4C sequences (Fig. 1).

That evolution of HIV in vivo can be discontinuous is shown by the failure to detect intermediate forms between the major sequence types, despite the fact that numerous base changes and more than one insertion or deletion event have occurred in the development of variant V4 and V5 sequences from the seroconversion type. Further evidence for the existence of hidden evolution is provided by the repeated observations in both the V4 and V5 regions that each succeeding sequence type is not obviously more related to those that come before or after it than they are to the sequence of the original infecting virus.

It has frequently been argued that sequence change in the *env* region may be an adaptive response by HIV to evade recognition by the immune system. Several studies have shown high rates of amino acid substitutions precisely in those areas of the immunodominant loop that are the targets of B-cell and T-cell recognition (1, 3, 12, 20, 36, 38, 44, 46) and in the equivalent region of the simian immunodeficiency virus genome of infection of rhesus monkeys (5). Indirect evidence for positive selection for sequence change in V3 is provided by a depressed synonymous-to-nonsynonymous ratio (K_s/K_a [21]) of nucleotide substitutions, significantly below 1, in the V3 loop region (36). In the current study, we have also found high rates of sequence change in these areas and could interpret the turnover of V3 sequence variants as a succession of escape mutants whose evolution is favored by a transient failure by the host immune system to neutralize the newly emergent forms. As was found in the V4 and V5 regions, succeeding V3 sequences are not necessarily direct derivatives of the previous V3 types. The predominant sequence in the 1989 RNA sample differs less from the seroconversion sequence than it does from the preceding variant (found at high frequency in the 1988 RNA sample and in the DNA population of the 1989 sample). As argued previously, whether the high rate of sequence change in the V4 and V5 regions is also a consequence of immune selection is not clear. Mouse antiserum to a peptide corresponding to the V5 region could neutralize HIV (15), although titers were lower than that of the antiserum raised against the V3 peptide, consistent with other peptide mapping studies (11, 17, 25, 28, 30, 32). However, the mature HIV gp120 protein is heavily glycosylated (19), and many of the potential sites for N-linked addition of carbohydrate are concentrated in the V4 and V5 hypervariable regions. These posttranslational modifications and long-range interactions with other regions of *env* on folding of the mature protein are likely to contribute to the formation of predominantly conformational epitopes in these regions.

Long-term persistence of seroconversion-type sequences. Many of the proviral sequences from p80 and p82 in samples taken several years after primary infection were identical to those detected at seroconversion. The absence of sequence change in the some of the most variable areas of the HIV genome is, at first sight, inconsistent with the generally high mutation rates associated with HIV replication (3, 33, 46). One explanation for complete absence of either silent or nonsilent changes over the entire V3-C2-V4-C3-V5 region is that the HIV proviral sequences detected in the PBMC samples in 1987 (p82) or 1989 (p80) that correspond to the seroconversion-type sequences have not replicated to any significant extent during the intervening years.

Supporting this hypothesis is the observation that HIV preferentially infects a long-lived cell subset of PBMCs in vivo. Almost all of the provirus detected in PBMCs is present in the $CD4^+$ lymphocyte fraction (35), of which the memory cell subset ($CD45RO^+$; $CD29^+$) appears to be preferentially infected in vivo by both HIV (34) and simian immunodeficiency virus (43). Consistent with their function in antigenic recall, it has been shown in humans (7) and by adoptive transfer in mice (16) that T memory cells can have essentially unlimited life spans relative to that of the host. Although HIV is normally considered cytopathic for T lymphocytes, a proportion of activated T lymphocytes may survive infection during the primary HIV infection and continue to circulate as differentiated memory cells with an unchanged proviral sequence. Thus, the persistence in p82 of V4B DNA sequences until 1988, when almost all RNA sequences were of the V4C type, may have been the consequence of long-term persistence of cells nonproductively infected in 1987.

There are many possible explanations for the proposed long-term survival of T cells infected at seroconversion and in subsequent years. Firstly, proviral sequences in those PBMCs that survive infection may contain inactivating mutations that prevent subsequent virus replication. High frequencies of defective proviral sequences have been reported to exist in vivo (26). However, using the limiting dilution PCR method that eliminates in vitro copying errors during

amplification (37), we have found an extremely low rate of inactivating substitutions in the *gag* and *env* regions (1 in over 40,000 bp sequenced) [3] and 3 in 110,000 bp [unpublished observations]). Furthermore, it has been shown that a high proportion of proviral sequences present in PBMCs can be activated in vitro to give replication-competent virus (4). Thus, defective viruses probably contribute little to persistent, nonlytic infection of lymphocytes.

An alternative explanation for the failure of HIV to kill the cell it infects is that the provirus may integrate into sites in the human genome that preclude or reduce the efficiency of cellular and viral mechanisms of transcription initiation, as has been described for other retroviruses (42).

Finally, the variants detected in the PBMC population may be replication competent and capable of activation but may contain mutations that make them less cytopathic for lymphocytes and allow the infected cells to survive. Supporting this latter possibility is the extensive in vitro evidence that isolates from HIV-infected individuals taken in early stages of infection are often noncytopathic for T lymphocytes, grow poorly in culture, and are incapable of any growth in T cell lines ("slow/low variants") (2, 6, 41). Virus variants in cells that have survived infection with HIV may therefore be a highly selected subset of the original infecting virus strain, whose noncytopathic (or defective) properties ensure their long-term survival without any need for continuous replication.

Origin of plasma and PBMC virus populations. We have provided extensive evidence for the existence of differences in the frequencies of different sequence types of virus present in plasma compared with those of proviruses present in HIV-infected PBMCs. Corroboration of the results obtained from sequencing (Fig. 3) was obtained by length analysis of amplified PCR products (Fig. 5 and Table 2), which discounted any effect of sampling error due to the small number of sequences. That there should be a difference between the two populations was unexpected, although it is not necessarily inconsistent with current theories of the pathogenesis of HIV (see below).

A consistent observation of this study was that each of the sequence types that initially appeared in the plasma RNA population eventually became the predominant PBMC sequence type. For this reason, the hypothesis that the unequal distribution of sequence types can be explained by their differing cell tropisms cannot be sustained. Similarly, comparisons of V3 proviral sequences in brain and spleen biopsy samples from three HIV-infected individuals has failed to reveal any systematic tissue-specific sequence differences (9). Only one of the three individuals showed major differences in the frequencies of distinct sequence types between the two tissues, and in the light of the data presented here, it is clearly possible that this difference was merely temporal.

The source of virus in the plasma could therefore be a subset of transcriptionally active $CD4^+$ lymphocytes, or virus could be secreted into the circulation by cells sequestered in solid tissue. It has been shown that plasma of both symptomatic and asymptomatic individuals is infectious (14), and thus infection of PBMCs may be a self-sustaining process. Infection, and continued sequence evolution of HIV, may indeed take place in peripheral $CD4^+$ lymphocytes.

Despite the high titers of infectivity of plasma in vitro, only a low proportion of T lymphocytes are infected in vivo (31, 35, 37), and an even lower proportion expresses detectable levels of virally encoded mRNA (13). Productive infection of T lymphocytes is thought to require T-cell activation by specific antigen or mitogen (39, 47, 48); thus, the observation that only 1 in 100 to 1 in 100,000 T lymphocytes contains provirus reflects the low frequency of activated cells in peripheral circulation. Although there is some evidence that the block to complete replication in nonactivated lymphocytes is at the level of integration and virus expression (39), it has been recently shown that virus replication may be prevented by incomplete reverse transcription of the incoming viral RNA (47). Furthermore, the truncated transcript is unstable and rapidly degraded, helping to explain the low frequency of provirus-bearing PBMCs in vivo.

According to the model advanced here, at any one time, proviral DNA sequences are composed of two distinct populations. Firstly, there are complete integrated copies of provirus in $CD45^+$ lymphocytes with no or minimal virus expression. Within the same sample, there would also be $CD4^+$ lymphocytes containing proviral DNA that were actively infected with HIV of the sequence types present in the plasma. The relative proportions of the two types of DNA would depend on the degree of infectivity of the plasma. We have previously shown that the amount of viral RNA present in plasma of HIV-infected individuals varies considerably, although there is a trend for symptomatic individuals to show higher concentrations than asymptomatic individuals (49). Similarly, the infectivity titers of plasma samples from patients with AIDS and AIDS-related complex were considerably higher than in those from patients with no evidence of clinical progression (14). Thus, in the early asymptomatic stages of infection, the majority of DNA sequences may remain of the seroconversion type, while, on progression, higher levels of infectious virus lead to increasing numbers of proviral sequences from secondary infection of lymphocytes, whose sequences would correspond to those of the plasma virus. It is notable that the apparent replacement of HIV sequences in PBMCs took place at a time when the proportion of infected cells was increasing (Table 1). Entirely consistent with the data given is the hypothesis that PBMCs containing the seroconversion type sequences remained in similar numbers for several years but were not detected after 1987, because they were numerically overtaken by the increasing numbers of PBMCs containing proviral sequences of the derived V4 and V5 sequence types.

The most direct test of whether plasma viral sequences are preferentially expressed in PBMCs is to compare the population of HIV mRNA sequences with those of PBMC provirus and plasma sequences. Unfortunately, it would not be possible to differentiate genuine mRNA sequences from viral RNA sequences that are present in the cytoplasm as a result of infection from plasma. According to the model advanced by Zack et al. (47), even nonactivated T cells may be susceptible to HIV attachment and entry; thus, within a PBMC sample, a large proportion of T lymphocytes may contain intact RNA templates from exogenous virus.

A major consequence of the difference in the compositions of the DNA and RNA populations is that sequential studies of virus evolution that are based on viral isolations from PBMC samples may be misleading. Because new sequence variants are initially more common in plasma than they are in PBMCs, isolations from the latter source may be composed predominantly or exclusively of previous virus types, possibly even of the seroconversion sequence. We are currently studying antigenic recognition of sequence-dependent epitopes in the V3 region to investigate the time course of development of specific immunity in p82, using oligopeptides

corresponding to each of the different RNA and DNA sequences obtained in this study.

ACKNOWLEDGMENTS

We are indebted to A. Cleland for technical assistance, to the staff of the Haemophilia Centre, Royal Infirmary of Edinburgh, for collection of patient samples, and to S. Rebus, D. Beatson, J. F. Peutherer, and M. Steel for providing background virological and immunological data on the two hemophiliac patients studied. We acknowledge the assistance of J. McKeating, J. O. Bishop, E. Holmes, and H. G. Watson for critical reading of the manuscript.

The work was supported by the Medical Research Council AIDS Directed Programme. L. Q. Zhang is supported by grants from the British Council and from the People's Republic of China.

REFERENCES

1. **Alizon, M., S. Wain-Hobson, L. Montagnier, and P. Sonigo.** 1986. Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. Cell **46**:63–74.
2. **Asjo, B., J. Albert, A. Karlsson, L. Morfeldt-Manson, G. Biberfeld, K. Lidman, and E. M. Fenyo.** 1986. Replicative properties of human immunodeficiency virus from patients with varying severity of HIV infection. Lancet **ii**:660–662.
3. **Balfe, P., P. Simmonds, C. A. Ludlam, J. O. Bishop, and A. J. L. Brown.** 1990. Concurrent evolution of human immunodeficiency virus type 1 in patients infected from the same source: rate of sequence change and low frequency of inactivating substitutions. J. Virol. **64**:6221–6233.
4. **Brinchmann, J. E., J. Albert, and F. Vartdal.** 1991. Few infected $CD4^+$ T cells but a high proportion of replication-competent provirus copies in asymptomatic human immunodeficiency virus type 1 infection. J. Virol. **65**:2019–2023.
5. **Burns, D. P. W., and R. C. Desrosiers.** 1991. Selection of genetic variants of simian immunodeficiency virus in persistently infected rhesus monkeys. J. Virol. **65**:1843–1854.
6. **Cheng-Mayer, C., D. Seto, M. Tateno, and J. A. Levy.** 1988. Biologic features of HIV-1 that correlate with virulence in the host. Science **240**:80–82.
7. **Darby, S. C., R. Doll, S. K. Gill, and P. G. Smith.** 1987. Long-term mortality after a single treatment course with x-rays in patients treated for ankylosing spondylitis. Br. J. Cancer **55**:179–190.
8. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the Vax. Nucleic Acids Res. **12**:387–395.
9. **Epstein, L. G., C. Kuiken, B. M. Blumberg, S. Hartman, L. R. Sharer, M. Clement, and J. Goudsmit.** 1991. HIV-1 V3 variation in brain and spleen of children with AIDS: tissue-specific evolution within host-determined quasispecies. Virology **180**: 583–590.
10. **Fitch, W. M., and E. Margoliash.** 1967. Construction of phylogenetic trees. Science **155**:279–284.
11. **Goudsmit, J., M. C. Debouck, R. H. Meloen, L. Smit, M. Bakker, D. M. Asher, A. V. Wolff, C. J. Gibbs, and D. Carleton.** 1987. Human immunodeficiency virus type 1 neutralization epitope with conserved architecture elicits early type-specific antibodies in experimentally infected chimpanzees. Proc. Natl. Acad. Sci. USA **85**:4478–4482.
12. **Hahn, B., G. M. Shaw, M. E. Taylor, R. R. Redfield, P. D. Markham, S. Z. Salahuddin, F. Wong-Staal, R. C. Gallo, E. S. Parks, and W. P. Parks.** 1986. Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. Science **232**:1548–1553.
13. **Harper, M. E., L. M. Marselle, R. C. Gallo, and F. Wong-Staal.** 1986. Detection of lymphocytes expressing T-lymphotropic virus type III in lymph nodes and peripheral blood from infected individuals by *in situ* hybridization. Proc. Natl. Acad. Sci. USA **83**:772–776.
14. **Ho, D. D., T. Mougdil, and M. Alam.** 1989. Quantitation of human immunodeficiency virus type 1 in the blood of infected individuals. N. Engl. J. Med. **321**:1621–1625.
15. **Ho, D. D., M. G. Sargadharan, M. S. Hirsch, R. T. Schooley, T. R. Rota, R. C. Kennedy, T. C. Chanh, and V. L. Sato.** 1987. Human immunodeficiency virus neutralizing antibodies recognize several conserved domains on the envelope glycoproteins. J. Virol. **61**:2024–2028.
16. **Jamieson, B. D., and R. Ahmed.** 1989. T cell memory. Long-term persistence of virus-specific cytotoxic T cells. J. Exp. Med. **169**:1993–2005.
17. **Javaherian, K., A. J. Langlois, C. McDanal, K. L. Ross, L. I. Eckler, C. L. Jellis, T. Profy, J. R. Rusche, D. P. Bolognesi, S. D. Putney, and J. Matthews.** Principal neutralizing domain of the human immunodeficiency virus type 1 envelope protein. Proc. Natl. Acad. Sci. USA **86**:6768–6772.
18. **Kimura, M.** 1980. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.
19. **Kozarsky, K., M. Penman, L. Basiripour, W. Haseltine, J. Sodroski, and M. Krieger.** 1989. Glycosylation and processing of the human immunodeficiency virus type 1 envelope protein. J. Acquired Immune Defic. Syndr. **2**:163–169.
20. **LaRosa, G. J., J. P. Davide, K. Weinhold, J. A. Waterbury, A. T. Profy, J. A. Lewis, A. J. Langlois, G. R. Dressman, R. N. Boswell, P. Shadduck, L. H. Holley, M. Karplus, D. P. Bolognesi, T. J. Matthews, E. A. Emini, and S. D. Putney.** 1990. Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. Science **249**:932–935.
21. **Li, W.-H., C.-I. Wu, and C.-C. Luo.** 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2**:150–174.
22. **Looney, D. J., A. G. Fisher, S. D. Putney, J. R. Rusche, R. R. Redfield, D. S. Burke, R. C. Gallo, and F. Wong-Staal.** 1988. Type-restricted neutralization of molecular clones of human immunodeficiency virus. Science **241**:357–359.
23. **Ludlam, C. A., J. Tucker, C. M. Steel, R. S. Tedder, R. Cheingsong-Popov, R. A. Weiss, D. B. McClelland, I. Philip, and R. J. Prescott.** 1985. Human T-lymphotropic virus type III (HTLV-III) infection in seronegative haemophiliacs after transfusion of factor VIII. Lancet **ii**:233–236.
24. **MacKeating, J. A., J. Gow, J. Goudsmit, L. H. Pearl, C. Mulder, and R. Weiss.** 1989. Characterisation of HIV-1 neutralisation escape mutants. AIDS **3**:777–784.
25. **Meloen, R. H., R. M. Liskamp, and J. Goudsmit.** 1989. Specificity and function of the individual amino acids of an important determinant of HIV-1 that induces neutralising antibody. J. Gen. Virol. **70**:1505–1512.
26. **Meyerhans, A., R. Cheynier, J. Albert, M. Seth, S. Kwok, J. Sninsky, L. Morfeld-Manson, B. Asjo, and S. Wain-Hobson.** 1989. Temporal fluctuations in HIV quasispecies *in vivo* are not reflected by sequential HIV isolations. Cell **58**:901–910.
27. **Modrow, S., B. H. Hahn, G. M. Shaw, R. C. Gallo, F. Wong-Staal, and H. Wolf.** 1986. Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: prediction of antigenic epitopes in conserved and variable regions. J. Virol. **61**:570–578.
28. **Neurath, A. R., N. Strick, and E. S. Y. Lee.** 1990. B cell epitope mapping of human immunodeficiency virus envelope glycoproteins with long (19- to 36-residue) synthetic peptides. J. Gen. Virol. **71**:85–95.
29. **Paabo, S., D. M. Irwin, and A. C. Wilson.** 1990. DNA damage promotes jumping between templates during enzymatic amplification. J. Biol. Chem. **265**:4718–4721.
30. **Palker, T. J., T. J. Matthews, M. E. Clark, G. J. Cianciole, R. R. Randall, A. J. Langlois, G. C. White, B. Safai, R. Snyderman, D. P. Bolognesi, and B. F. Haynes.** 1987. A conserved region at the COOH terminus of human immunodeficiency virus gp120 envelope protein contains an immunodominant epitope. Proc. Natl. Acad. Sci. USA **84**:2479–2483.
31. **Psallidopoulos, M. C., S. M. Schnittman, L. M. Thompson III, M. Baseler, A. S. Fauci, H. C. Lane, and N. P. Salzman.** 1989. Integrated proviral human immunodeficiency virus type 1 is present in CD4+ peripheral blood lymphocytes in healthy seropositive individuals. J. Virol. **63**:4626–4631.
32. **Rusche, J. R., K. Javaherian, C. McDanal, J. Petro, D. L. Lynn,**

R. Grimaila, A. Langlois, R. C. Gallo, L. O. Arthur, P. J. Fischinger, D. P. Bolognesi, S. D. Putney, and T. J. Matthews. 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. USA **85**:3198–3202.

33. **Saag, M. S., B. H. Hahn, J. Gibbons, Y. Li, E. S. Parks, W. P. Parks, and G. M. Shaw.** 1988. Extensive variation of human immunodeficiency virus type-1 *in vivo*. Nature (London) **334**: 440–444.
34. **Schnittman, S. M., H. C. Lane, J. Greenhouse, J. S. Justement, M. Baseler, and A. S. Fauci.** 1990. Preferential infection of CD4+ memory T cells by human immunodeficiency virus type 1: evidence for a role in the selective T-cell functional defects observed in infected individuals. Proc. Natl. Acad. Sci. USA **87**:6058–6062.
35. **Schnittman, S. M., M. C. Psallidopoulos, H. C. Lane, L. Thompson, M. Baseler, F. Massari, C. H. Fox, N. P. Salzman, and A. S. Fauci.** 1989. The reservoir for HIV-1 in human peripheral blood is a T cell that maintains expression of CD4. Science **245**:305–308.
36. **Simmonds, P., P. Balfe, C. A. Ludlam, J. O. Bishop, and A. J. L. Brown.** 1990. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. J. Virol. **64**:5840–5850.
37. **Simmonds, P., P. Balfe, J. F. Peutherer, C. A. Ludlam, J. O. Bishop, and A. J. L. 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.
38. **Starcich, B. R., B. H. Hahn, G. M. Shaw, P. D. McNeely, S. Modrow, H. Wolf, E. S. Parks, W. P. Parks, S. F. Josephs, R. C. Gallo, and F. Wong-Staal.** 1986. Identification and characterization of conserved and variable regions in the envelope gene of HTLVIII/LAV, the retrovirus of AIDS. Cell **45**:637–648.
39. **Stevenson, M., T. L. Stanwick, M. P. Dempsey, and C. A. Lamonica.** 1990. HIV-1 replication is controlled at the level of T cell activation and proviral integration. EMBO J. **9**:1551–1560.
40. **Takahashi, H., S. Merli, S. D. Putney, R. Houghten, B. Moss, R. N. Germain, and J. A. Berzofsky.** 1989. A single amino acid interchange yields reciprocal CTL specificities for HIV-1 gp160. Science **246**:118–121.
41. **Tersmette, M., R. E. Y. de Goede, B. J. M. Al, I. N. Winkel, R. A. Gruters, H. T. Cuypers, H. G. Huisman, and F. Miedema.** 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.
42. **Varmus, H., and R. Swanstrom.** 1985. Replication of retroviruses, p. 75–134. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
43. **Willerford, D. M., M. J. Gale, R. E. Benveniste, E. A. Clarke, and W. M. Gallatin.** 1990. Simian immunodeficiency virus is restricted to a subset of blood CD4+ lymphocytes that includes memory cells. J. Immunol. **144**:3779–3783.
44. **Willey, R. L., R. A. Rutledge, S. Dias, T. Folks, T. Theodore, C. E. Buckler, and M. A. Martin.** 1986. Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency virus syndrome retrovirus. Proc. Natl. Acad. Sci. USA **83**:5038–5042.
45. **Williams, P., P. Simmonds, P. L. Yap, P. Balfe, J. Bishop, R. Brettle, R. Hague, D. Hargreaves, J. Inglis, A. J. L. Brown, J. Peutherer, S. Rebus, and J. Mok.** 1990. The polymerase chain reaction in the diagnosis of vertically transmitted HIV infection. AIDS **4**:393–398.
46. **Wolfs, T. F. W., J. de Jong, H. van der Berg, J. M. G. H. Tunagel, W. J. A. Krone, and J. Goudsmit.** 1990. Evolution of sequences encoding the principal neutralization epitope of HIV-1 is host-dependent, rapid and continuous. Proc. Natl. Acad. Sci. USA **87**:9928–9942.
47. **Zack, J. A., S. J. Arrigo, S. R. Weitsman, A. S. Go, A. Haislip, and I. S. Y. Chen.** 1990. HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile, latent viral structure. Cell **61**:213–222.
48. **Zagury, D., J. Bernard, R. Leonard, R. Cheynier, M. Feldman, P. S. Sarin, and R. C. Gallo.** 1986. Long term cultures of HTLV-III-infected T cells: a model of cytopathology of T-cell depletion in AIDS. Science **231**:850–853.
49. **Zhang, L. Q., P. Simmonds, C. A. Ludlam, and A. J. L. Brown.** 1991. Detection, quantitation and sequencing of HIV-1 virus from the plasma of seropositive individuals and from factor VIII concentrates. AIDS **5**:675–681.

Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 4835–4839, June 1992
Evolution

Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient

(V3 loop evolution/phylogeny/escape mutants/neutralization)

EDWARD C. HOLMES*†, LIN QI ZHANG*, PETER SIMMONDS‡, CHRISTOPHER A. LUDLAM§, AND ANDREW J. LEIGH BROWN*

*Division of Biological Sciences, Institute of Cell, Animal and Population Biology, Crew Building, West Mains Road, Edinburgh EH9 3JN, Scotland; ‡Department of Medical Microbiology, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, Scotland; and §Department of Hematology, Royal Infirmary of Edinburgh, Lauriston Place, Edinburgh EH3 9YW, Scotland

Communicated by J. Maynard Smith, February 3, 1992

ABSTRACT **In an investigation of the evolution of the third hypervariable loop of gp120 (V3), the principal neutralization determinant of human immunodeficiency virus type 1, we have analyzed 89 V3 sequences of plasma viral RNA purified from peripheral blood samples donated over 7 years by an infected hemophiliac. Considerable sequence diversity in the V3 region was found at all time points after seroconversion. Phylogenetic analysis revealed that an important diversification had occurred by 3 years postinfection and that, subsequently, most sequences could be allocated to either one of two major lineages that persisted throughout the remainder of the infection. Rapid changes in frequency of the most common sequences and the observation that the same hexapeptide motif (GPGSAV) at the crown of the V3 loop has evolved convergently provide strong evidence that selective processes determine the evolutionary fate of sequence variants in this region.**

Analyses of nucleic acid sequences from human immunodeficiency virus type 1 (HIV-1) have revealed a large amount of variation both between and within patients (1–10). This variation is not evenly distributed across the HIV-1 genome; rates of nucleotide substitution are particularly high in hypervariable regions (V1–V5) of the *env* gene (11) and are lower in *pol* and the p24 coding region of *gag* (1, 11, 12). This suggests that the levels of sequence variation are not simply reflections of the intrinsic mutation rate.

There are good grounds for expecting that natural selection will be important in determining HIV-1 variability. This may be manifest as selective constraint against variation in amino acid sequence, as in the *pol* gene and the p24 region of *gag* (11, 12). On the other hand, studies of two other lentiviruses, equine infectious anemia virus and visna virus, suggest that much of the variability seen in *env* may be adaptive. In both of these viruses, antigenically distinct isolates were found to arise progressively during infection, always appearing before their associated neutralizing antibodies (refs. 13 and 14 and references therein). Differences between isolates could be assigned to mutations clustered in the *env* genes, the majority of which resulted in amino acid replacement. Thus, it was proposed that viral mutants arise that can "escape" from recognition by neutralizing antibodies. This change in the genetic composition of the viral population, referred to as "antigenic drift" although it is a selective process, is generally considered to be one of the principal mechanisms by which lentiviruses evade host immune systems (reviewed in ref. 15). However, high levels of variability are also seen in the *nef* and *tat* genes (2, 6).

The *env* gene encodes the principal neutralization determinant of HIV-1 (5, 16), which has been mapped to an ≈35-amino acid residue disulfide-bonded loop structure in the third hypervariable region of surface envelope glycoprotein gp120 (V3) (16–18). There have been a number of attempts to assess *in vitro* the fitness of sequential V3 loop isolates by serological techniques (19) and to examine patterns of sequence change *in vivo* (9, 10). Simmonds *et al.* (20) analyzed sequence change in the adjacent V4 and V5 hypervariable regions of *env* over a 5-year period after seroconversion in an asymptomatic hemophiliac (9). It was concluded that analysis of HIV sequence change in *env* within a patient based only on the lymphocyte-associated proviral population could be misleading because of significant differences in the frequency and persistence of sequence variants compared with the plasma (considered to be recently replicated) virus population. Therefore, analysis of variation in the V3 region presented here is based on sequences detected in plasma virus particles.

To assess the role and nature of selection acting on V3 sequences, we have carried out a detailed analysis of the process of viral evolution in this region during the course of an infection.

MATERIALS AND METHODS

Nucleotide Sequences. Eighty-nine viral RNA sequences (plasma derived) of ≈240 base pairs were obtained directly from nested PCR amplified single molecules from an Edinburgh hemophiliac patient (p82) infected from a locally prepared batch of factor VIII in 1984 (1). Sequences were obtained in varying numbers (n) from seroconversion [year 0 (1984, n = 12); year 3 (1987, n = 15); year 4 (1988, n = 11); year 5 (1989, n = 23); year 6 (1990, n = 15); year 7 (1991, n = 11) postinfection]. These sequences have been assigned GenBank accession numbers M84240–M84317. Methods used for extraction and PCR amplification of single cDNA molecules are described by Zhang *et al.* (21) and those for nested PCR amplification and direct sequencing are described by Simmonds *et al.* (20). Clinical data for this patient are given in ref. 9. The patient has been asymptomatic throughout the period of study and has never received antiviral therapy.

Phylogenetic Analysis. V3 sequences from six isolates from the USA [HIV_{RF}, HIV_{WMJ2}, HIV_{MN}, HIV_{SF2}, HIV_{PV22}, and $HIV_{HTLVIIIB}$ (clone HXB2)] were used as outgroups (all were taken from the Los Alamos Human Retroviruses and AIDS

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HIV-1, human immunodeficiency virus type 1; V3, third hypervariable region of surface envelope glycoprotein gp120.

†To whom reprint requests should be addressed.

data base). Sequences were aligned by hand, as the only length variation was the dipeptide insertion of QR between residues 309 and 310 in isolates HIV_{PV22} and $HIV_{HTLVIIIB}$ (HXB2).

All phylogenetic analysis programs used were taken from the PHYLIP package (version 3.4) provided by J. Felsenstein (Department of Genetics, University of Washington). The principal program used was DNAML, which implements a maximum likelihood method. Global branch swapping was used, as this increases the proportion of trees searched. The default settings were used for the other options. We also used the neighbor-joining distance matrix method of Saitou and Nei (22) as implemented in the PHYLIP program NEIGHBOR. Distances were estimated by using the same evolutionary model as underlies the DNAML program by the program DNADIST.

RESULTS

Phylogenetic Analysis of the V3 Region. Phylogenetic relationships can be inferred with greatest confidence when there is no homoplasy (convergence) in the data. As selection can cause convergence even at the sequence level (23), we divided the region under study into three sections—the 35-amino acid V3 loop itself and two flanking regions (a 19-residue 5′ region and a 23-residue 3′ region)—and carried out all phylogenetic analyses on nucleotide sequences. Trees were inferred from different combinations of these regions taken separately, as well as from the entire sequence, in order not to bias the analysis toward any spurious phylogenetic patterns found only in the loop.

An unusual feature of the data was that, whereas in most phylogenetic analyses sequences are contemporaneous, sequences in this data set were sampled from different time points during the evolution of the viral population within an infected patient. Therefore, trees were also inferred for each year separately, in addition to analyses of the whole data set.

Relationships of V3 Sequences. The most striking and consistent result from the phylogenetic analyses of the nucleotide sequence data was that a major division into two distinct lineages had occurred by 1987. This is depicted in the global maximum likelihood tree for all distinct V3 sequences of plasma virions from p82 (Fig. 1). For reasons of clarity, only branches connecting the 46 p82 sequences that were separated by branch lengths found to be significantly different from 0 under the DNAML model are shown. Individual sequences are represented by a dot. A single USA outgroup sequence, $HIV_{HTLVIIIB}$ clone HXB2, is also included. The subdivision into two major descendent lineages, D and E (see below), is indicated. Although not drawn so, this tree may be rooted by HXB2 (as labeled), in which case evolution may be thought of as running from left to right. Twelve identical V3 sequences were obtained at seroconversion. All trees showed this sequence to be the ancestor of all others in that it is closest to the root (as labeled in Fig. 1). Finally, all branch lengths are drawn to scale, which permits an assessment of the relative amounts of evolutionary change along different lineages.

Other, smaller, bifurcations were also consistently observed and these, like the major subdivision, were found to accord with the pattern of amino acid replacements in the V3 loop (see below). Conversely, a few sequences were more difficult to place and frequently changed position in different analyses. These were mainly sequences from year 3, a sample in which most sequences had diverged to a similar extent (see below). By year 4, however, the major lineages were clearly distinguished.

Evolution of the V3 Loop. Having established the phylogenetic relationships between sequences from the V3 region from this patient, we wished to determine, because of the functional importance of the V3 loop itself, whether the lineages we have identified correspond to changes in the antigenic structure of the loop. In the entire data set, there were a total of 24 different V3 loop amino acid sequences. These sequences are listed in Table 1, together with the frequencies at which they are found in each year. Each can be seen as having one or more amino acid differences from that found at seroconversion (designated sequence A). If we superimpose the V3 loop amino acid sequences onto the trees deduced from the nucleotide sequences (Fig. 1, for example), we obtain the evolutionary "framework" depicted in Fig. 2. Distinct evolutionary lineages of V3 loop sequences have been identified and assigned a letter (A–F) and each different sequence within these lineages has been assigned a number (Fig. 2 and Table 1). The proposed relationships and the directions of changes between the sequences of a lineage are indicated by arrows (Fig. 2). Sequences which have persisted through time points are indicated by dashed arrows. Relationships between some lineages were difficult to assign, as different analyses produced different phylogenies, so affiliations between lineages have only been indicated when they were unequivocal. For example, the relationships between lineages B, E, and F are uncertain and the symbol ? that connects them is used to signify this uncertainty. Finally, unlike most molecular phylogenies, real ancestors may be

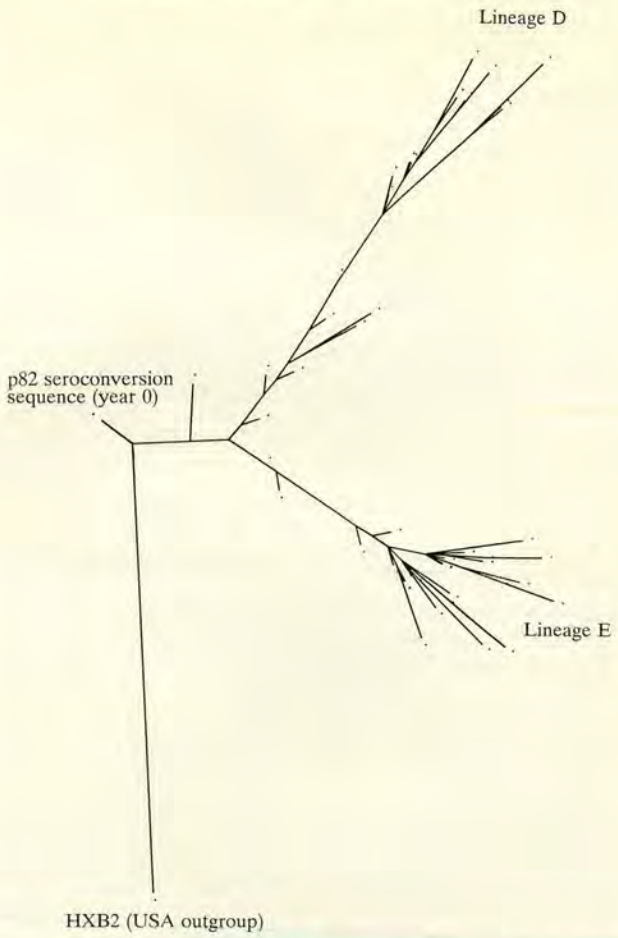


FIG. 1. Maximum likelihood tree depicting evolutionary relationships among V3 sequences obtained from plasma virus from p82. This tree may be rooted by the HXB2 outgroup sequence. Given this rooting, the first p82 sequence to appear is that found at seroconversion, as indicated. The major evolutionary subdivision is clearly visible and the two major descendent lineages, D and E, are also labeled. Dots at the tips of the lineages represent individual sequences; all branch lengths are drawn to scale.

Table 1. The 24 V3 loop amino acid sequences and their sample frequencies in the plasma

Lineage	Sequence	Frequency in year 0	3	4	5	6	7
	296 CTRPNNNTRKSIHIGPGRAFYTTGEIIGDIRQAHC 330						
A		1.000					
B	.....D.....		0.067				
C1	.....P.....D.....		0.067				
C2	.....P.....Q.....		0.267				
C3	.....G.....Q.....		0.267				
C4	.....P.....D.T.....		0.067				
C5	.....D.T.....		0.067				
D1	.....V....Q.....			0.091			
D2	.....R.....V...EQ...N.....			0.455			
D3	.....G.....V...EQ...N.....			0.091			
D4	.....R.Y.....V...EQ...N.....				0.087		
D5	.....R.Y....S.V...EQ...N.....				0.043		
D6	.....R.Y.....V...DQ.....					0.067	
D7	.....R.Y.....V...DQ...N.....					0.200	0.077
D8	.....Y...R.G.....SV..AEQ...N.....					0.200	0.077
E1	.....G.....S...A..D.....		0.067		0.043	0.333	0.769
E2	.....G.....S...A..G.....			0.182	0.696		
E3	.....G.....S...A..R.....			0.091			
E4	.....G.....S.V.A..G.....				0.043		
E5	.....G.....S...A..G...N.....				0.087		
E6	.....G.....S.V.A..D.....					0.067	
E7	.....G.....A..D.....					0.067	
E8	.....G.....V....D.....					0.067	
F	.....G.....		0.133	0.091			0.077
	Total	12	15	11	23	15	11

Amino acid sequences are listed according to their evolutionary lineage (A–F). Only residues that differ from those in sequence A (detected at seroconversion) are shown, with a dot denoting identical residues. Amino acid positions are numbered according to Wolfs *et al.* (10). The frequencies of each sequence in each year from which a sample was available are presented, with the total number of sequences obtained given at the bottom. Blank space indicates that the sequence was not detected in that particular year.

present in the data and the framework expresses the postulated ancestor–descendent relationships.

Many of the evolutionary patterns observed at the nucleotide level, and depicted in Fig. 1, correspond to amino acid changes in the loop itself. From Figs. 1 and 2 and Table 1, it is clear that evolutionary lineages D and E dominate the data set. Lineage D can be distinguished by 2 amino acid replacements from its apparent ancestor—C2. These are valine for phenylalanine at position 315 and glutamine for glutamic acid at position 320. A number of other replacements are also found in most of the lineage D variants, such as aspartic acid to asparagine (N) at position 324. The relationship of the lineage E sequences to each other is characterized by a serine to glycine substitution at position 306. Most members of this lineage also have a serine at position 313 and an alanine at position 317.

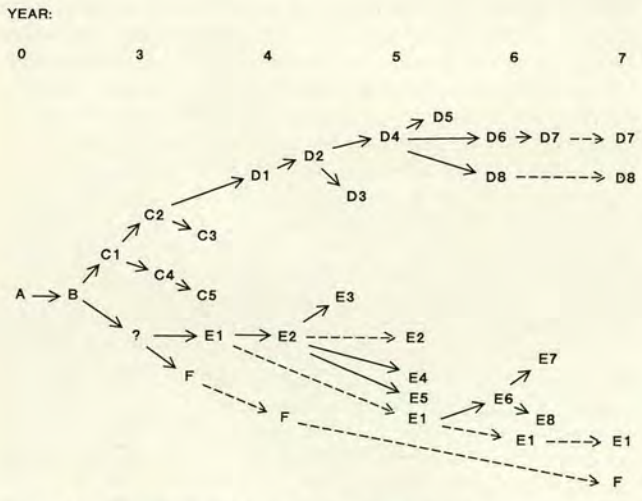


FIG. 2. Evolutionary framework relating the 24 different amino acid sequences found in the V3 loop. Distinct evolutionary lineages are designated by letters A–F and sequences within lineages are designated by numbers. Proposed relationships are indicated by arrows. Lineages that persist through years are indicated by dashed arrows. Time scale is given along the top.

Comparison of lineages D and E reveals some interesting features in the evolution of the 6 amino acids that correspond to the crown of the V3 loop (10). In the seroconversion sample, these were GPGRAF (Table 1). The same motif has also been found in 146/256 (60%) isolates of HIV-1 from the USA examined by La Rosa *et al.* (5). Lineages D and E have undergone a number of changes in this sequence; GPGRAV is most frequently found in lineage D [although only 12 times (5%) in the set of isolates from the USA (5)], while GPGSAF [not reported by La Rosa *et al.* (5)] characterizes the early evolution of lineage E. Remarkably, both lineages acquired the motif GPGSAV independently.

As already noted, the relationships of lineages found in year 3 (B, C, E, and F) and of the sequences within the C lineage are harder to interpret. All analyses suggested that sequence B, characterized by a single amino acid change (glutamic acid to aspartic acid at position 320) from the seroconversion sequence A and found only once in the data set, is the ancestor of all later variants. The similarity between sequences A and B, in the face of the diversity found in year 3, suggests that B probably arose early in infection.

All members of the C lineage can be derived from sequence B. C4 and C5 are grouped through their possession of a threonine at position 322, while the relationship between C2 and C3 is suggested by their possession of a glutamine at position 320. The remaining intersequence relationships are even harder to define but, as depicted (Fig. 2), it seems most likely that C1 is the ancestor of all others.

Changes in Sequence Frequency. The frequencies at which the sequences are found in the plasma are listed in Table 1. This table highlights the dramatic changes in frequency of the most abundant sequences. The most notable examples are provided by sequence A, the only sequence found at seroconversion but not seen subsequently; E2, which constituted almost 70% of the sample in year 5 but was not found in any later samples; and D2, which made up 45% of the sample in year 4 but was not found afterward.

It is unlikely that these changes in frequency are simply sampling artifacts for two reasons. First, we have shown in a study of the V4 and V5 sequences from the same patient (9) that the sample frequencies estimated from nucleotide sequence data (where n was $\approx$10) accorded well with the relative frequencies of length variants observed by scanning densitometry in much larger samples of the same material (where n, the input copy number, was $\approx$100). Second, random sampling would affect most strongly those variants found at low frequency. The striking feature of the evolution described here (Table 1) is that the most dramatic changes involve those sequences that reach high frequency.

Convergent Evolution. One of the most unexpected observations stemming from the phylogenetic analysis of V3 sequences was that the same amino acid has been fixed independently at the same site in more than one sequence. The most spectacular examples involve the hexapeptide at the crown of the V3 loop. The evolution of variants of this motif, as inferred from the evolutionary framework (Fig. 2), is shown in Fig. 3. We infer the motif found at seroconversion (GPGRAF) to be the ancestor of all later motifs, including GPGSAV (indicated by an asterisk in Fig. 3), which appears to have evolved independently three times within p82—in sequences D5, E4, and E6—and twice in 1989 alone. Other instances of convergent evolution in this motif are GPGRAV, found in most of the D lineage sequences and in E8, and the original GPGRAF motif, which reappears in E7. Convergent evolution in this hexapeptide is evidently frequent. In a second hemophiliac patient from Edinburgh also infected in 1984 with a virus of the same sequence (9), the independent evolution of the GPGSAV motif has also been observed (data not shown), making four times in all. Further convergence occurs outside this motif; the glycine at position 308 appears to have arisen twice (in D3 and the E lineage) as does the glycine at position 320 (in three members of the E lineage—E2, E4, and E5—as well as in genotype F) and the asparagine at position 324 (in the D lineage and E5). Overall, this level of convergence seems to be most compatible with the interpretation that sequence change in this region represents adaptive evolution (23).

DISCUSSION

We have used a number of different methods of phylogenetic inference. This is particularly necessary when natural selection may be important. From simulation studies, it has become clear that maximum likelihood inference (as implemented in DNAML) performs extremely well, in terms of the proportion of times it retrieves the correct tree, under a neutral (Poisson based) model of molecular evolution (24). It is unclear, however, how any method fares under evolutionary processes that deviate from neutrality, and thus it is probably best not to rely solely on one.

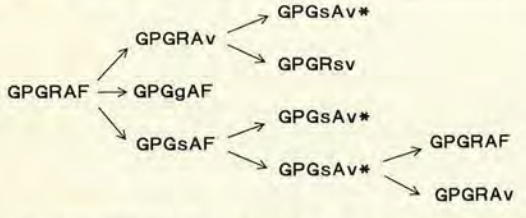


FIG. 3. Evolutionary relationships between the different hexapeptide motifs found at the crown of the V3 loop. The motif found at seroconversion (year 0) is GPGRAF and, as this is the inferred ancestor of all later motifs, is written in capital letters. From this, five different motifs evolve during the course of the infection, with the amino acids that differ from that found at seroconversion written in lowercase letters. Convergent evolution of the motif GPGSAV is denoted by an asterisk.

Evolution of HIV Sequences in p82. During the infection of p82 there has been an evolutionary diversification of V3 sequences. This diversification can be assigned to evolution along a number of major lineages, themselves characterized by successive substitutions at residues of (presumed) antigenic importance. Selection for viral mutants that can escape from the neutralizing capability of host antibodies is likely to be the principal mechanism that drives this evolutionary change (25). The rapid change in the composition of the V3 sequence population—the rise and fall of variants—is a consequence of this continual process of neutralization–escape. This is also reflected as alterations in the relative prominence of different lineages, specifically D and E, in samples from successive years. Thus, antigenic drift, instead of being the sequential replacement of one antigenically distinct variant by another, may involve a complex interaction between the different, and competing, evolutionary lineages present in the plasma.

The rapid changes in the plasma population also suggest that antigenic evolution has an important frequency-dependent element. Standard host–parasite models (26) suggest that the higher the frequency a viral variant reaches in the plasma, the higher the probability of its recognition and neutralization. Table 1 shows that in each year it is generally the most frequent sequence that appears to be neutralized (i.e., shows the greatest reduction in frequency). Consequently, variants found at low frequency will have a greater selective advantage and will increase in frequency until they too are countered with an effective immune response.

The evolution of the viral population is made more complicated by the fact that selection pressures may not be constant during the course of the infection. Nowak *et al.* (27, 28) suggest that the move from the asymptomatic to the symptomatic stage of infection is triggered by the loss of a specific immune response that contains a highly diverse viral population during the early stages and that, in AIDS, new antigenic variants may no longer be favored. It is interesting to note that no new V3 amino acid sequences were found in year 7—at which time the $CD4^+$ cell count of p82 had fallen below 200 (9), although the patient remained asymptomatic.

Constraints on V3 Loop Sequences. We interpret the observation of extensive convergent evolution to be due to the interplay between selection for variability and for conservation. Specifically, although there is positive selection for replacement of amino acids that remove B-cell (or T-cell) epitopes once these are recognized, there is also a selective constraint as to which amino acids are functionally viable within this region. Interestingly, many of the differences between the V3 loop sequences from p82 only involve a limited number of amino acid replacements. This is most clearly seen in the case of sequence E1, which replaces E2 as the most frequent in the population, although distinguished by only a single amino acid change (glycine to aspartic acid at position 320; Table 1). It is also possible that there are conformational changes elsewhere in the protein, which may compensate for changes in the loop itself.

Finally, the high proportion of nucleotide substitutions that lead to amino acid replacements (K_a) compared to silent substitutions (K_s) (9) is also consistent with the action of selection (data not shown).

Rate of Sequence Evolution. Several studies, including our own, have attempted to estimate a mean rate of sequence

evolution for HIV-1 (1, 4, 10, 11, 29). If, as we suggest, natural selection is the major mechanism of sequence change in this region of the HIV-1 genome, then any attempt to estimate a mean rate of evolution within a patient is likely to be misleading. In p82, the mean nucleotide distance between samples from years 5 and 6 was actually greater than that between years 4 and 6 (data not shown), and similar phenomena have been observed before (4, 10). This can now be seen to arise from the alternation between the two predominant evolutionary lineages in p82. Even within a lineage, we observe substantial variation in the rate of change. The concept of a mean rate of sequence evolution is almost meaningless in the context of a region such as V3, which is so strongly influenced by selection.

We thank the staff of the Hemophilia Centre (Royal Infirmary of Edinburgh) for taking samples, Sarah Ashelford for assisting with the sequencing, and two referees for helpful comments. This work was supported by the AIDS Directed Programme and the Committee on Epidemiological Studies of AIDS of the Medical Research Council.

1. Balfe, P., Simmonds, P., Ludlam, C. A., Bishop, J. O. & Leigh Brown, A. J. (1990) *J. Virol.* **64,** 6221–6233.
2. Delassus, S., Cheynier, R. & Wain-Hobson, S. (1991) *J. Virol.* **65,** 225–231.
3. Epstein, L. G., Kuiken, C., Blumberg, B. M., Hartman, S., Sharer, L. R., Clement, M. & Goudsmit, J. (1991) *Virology* **180,** 583–590.
4. Hahn, B., Shaw, G. M., Taylor, M. E., Redfield, R. R., Markham, P. D., Salahuddin, S. Z., Wong-Staal, F., Gallo, R. C., Parks, E. S. & Parks, W. P. (1986) *Science* **232,** 1548–1553.
5. La Rosa, G. J., Davide, J. P., Weinhold, K., Waterbury, J. A., Profy, A. T., Lewis, J. A., Langlois, A. J., Dreesman, G. R., Boswell, N., Shadduck, P., Holley, L. H., Karplus, M., Bolognesi, D. P., Matthews, T. J., Emini, E. A. & Putney, S. D. (1990) *Science* **249,** 932–935.
6. Meyerhans, A., Cheynier, R., Albert, J., Seth, M., Kwok, S., Sninsky, J., Morfeldt-Manson, L., Asjo, B. & Wain-Hobson, S. (1989) *Cell* **58,** 901–910.
7. Saag, M. S., Hahn, B. H., Gibbons, Y., Li, Y., Parks, E. S., Parks, W. P. & Shaw, G. M. (1988) *Nature (London)* **344,** 440–444.
8. Simmonds, P., Balfe, P., Ludlam, C. A., Bishop, J. O. & Leigh Brown, A. J. (1990) *J. Virol.* **64,** 5840–5850.
9. Simmonds, P., Zhang, L. Q., McOmish, F., Balfe, P., Ludlam, C. A. & Leigh Brown, A. J. (1991) *J. Virol.* **65,** 6266–6276.
10. Wolfs, T. F. W., De Jong, J.-J., Van Den Berg, H., Tunagel, J. M. G. H., Krone, W. J. A. & Goudsmit, J. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 9938–9942.
11. Li, W.-H., Tanimura, M. & Sharp, P. (1988) *Mol. Biol. Evol.* **5,** 313–330.
12. Leigh Brown, A. J. & Moynaghan, P. (1988) *AIDS Res. Hum. Retroviruses* **4,** 399–407.
13. Scott, J. V., Stowring, L., Hasse, A. T., Narayan, O. & Vigne, R. (1979) *Cell* **18,** 321–327.
14. Salinovich, O., Payne, S., Montelaro, R., Hussain, K., Issel, C. & Schnorr, K. (1986) *J. Virol.* **57,** 71–80.
15. Clements, J. E., Gdovin, S. L., Montelaro, R. C. & Narayan, O. (1988) *Annu. Rev. Immunol.* **6,** 139–159.
16. Javaherian, K., Langlois, A. J., McDanal, C., Ross, K. L., Eckler, L. I., Jellis, C. L., Profy, A. T., Rusche, J. R., Bolognesi, D. P., Putney, S. D. & Matthews, T. J. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 6768–6772.
17. Modrow, S., Hahn, B. H., Shaw, G., Gallo, R. C., Wong-Staal, F. & Wolf, H. (1987) *J. Virol.* **61,** 570–578.
18. Leonard, C. K., Spellman, M. W., Riddle, L., Harris, R. J., Thomas, J. N. & Gregory, T. J. (1990) *J. Biol. Chem.* **265,** 10373–10382.
19. McKeating, J. A., Gow, J., Goudsmit, J., Pearl, L. H., Mulder, C. & Weiss, R. (1989) *AIDS* **3,** 777–784.
20. Simmonds, P., Balfe, P., Peutherer, J. F., Ludlam, C. A., Bishop, J. O. & Leigh Brown, A. J. (1990) *J. Virol.* **64,** 864–872.
21. Zhang, L. Q., Simmonds, P., Ludlam, C. A. & Leigh Brown, A. J. (1991) *AIDS* **5,** 675–681.
22. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425.
23. Stewart, C. B., Schilling, J. W. & Wilson, A. C. (1987) *Nature (London)* **330,** 401–404.
24. Saitou, N. & Imanishi, T. (1989) *Mol. Biol. Evol.* **6,** 514–525.
25. Wolfs, T. F. W., Zwart, G., Bakker, M., Valk, M., Kuiken, C. L. & Goudsmit, J. (1991) *Virology* **185,** 195–205.
26. Anderson, R. M. & May, R. M. (1982) *Parasitology* **85,** 411–426.
27. Nowak, M. A., May, R. M. & Anderson, R. M. (1990) *AIDS* **4,** 1095–1103.
28. Nowak, M. A., Anderson, R. M., McLean, A. R., Wolfs, T. F. W., Goudsmit, J. & May, R. M. (1991) *Science* **254,** 963–969.
29. Yokoyama, S. & Gojobori, T. (1987) *J. Mol. Evol.* **24,** 330–336.