SEQUENCE ANALYSIS OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1:

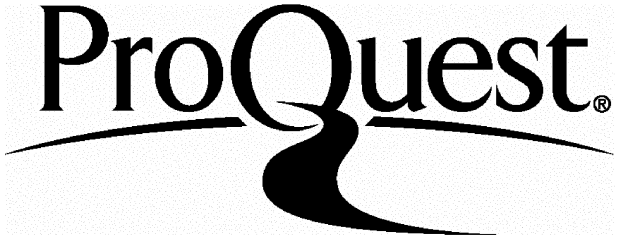
A CROSS-SECTIONAL AND LONGITUDINAL ANALYSIS.

By

Mounir Ait-Khaled

Department of Virology
Royal Free Hospital School of Medicine
University of London

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy for the University of London, 1996.



08903

ABSTRACT

The replication of HIV-1 and the turnover of CD4 cells *in vivo* is a highly dynamic process. This environment of continuous *de novo* virus infection and extensive replication provides an ideal setting for the generation of HIV-1 variants. The objectives of this thesis were to perform a cross-sectional and longitudinal sequence analysis of HIV-1 in blood and multiple tissues of HIV-1 infected individuals. Five to 20 clones of Polymerase Chain Reaction (PCR) amplicons were sequenced and their relationships investigated using phylogenetic methods. An analysis of V3 sequences from a patient at the time of seroconversion revealed a homogeneous quasispecies which contrasted with the high diversity observed in the V3 quasispecies from a patient with AIDS. Phylogenetic analysis of HIV-LTR variants from LN and blood from 4 patients revealed 2 distinct patterns. First, an independent clustering of LN and PBMC variants, indicating a compartmentalisation of the two quasispecies for the patient with an intact lymph node. Second, non-polarised trees with variants from either LN or PBMC present in common branches for the patients with a disrupted LN architecture. These results suggest that the lymph nodes may be important in the sequestration/evolution of distinct HIV-1 variants. A similar phylogenetic analysis has been carried out on LTR variants from multiple post-mortem samples (spleen, lung, spinal cord, ganglion, lymph node and blood) of a patient who died with AIDS. The results revealed a genetically distinct LTR quasispecies in the nervous tissues compared to that present in the other tissues. The nervous tissue quasispecies was characterised by 7 mutations in the normally conserved TAR region which would be expected to abrogate *tat* transactivation. However, this quasispecies showed the maintenance of prototypic NF-κB sites which would allow TAR independent κB *tat* transactivation to occur. These data provide evidence for a role of the LTR in the adaptation of HIV-1 to the nervous tissue environment by optimising the use of specific cellular transcription factor(s).

ACKNOWLEDGEMENTS

I am indebted to Doctor Vincent Emery for his guidance, supervision, and enthusiastic support.

I am grateful to Professsor Paul Griffiths and Doctor Deenan Pillay for their advice and critical discussions.

Credit is due to the generosity of the patients, and to the Royal Free Hospital HIV medical team who care for them.

I would like to thank all the members of the Virology Department at the Royal Free Hospital for their help, and particularly Elaine Lipscombe, Mike Kidd and Mark Grzywacz for enhancing my computing skills with great kindness and patience.

I thank Doctor G. Lang for her help with the band shift assays.

Financial support from the Algerian government is acknowledged.

Finally, and by no means least, I am grateful to the few close friends who endured with me the difficult final moments of my postgraduate life, particularly, I am indebted to Kyriacos Mitrophanous for his valuable friendship, and his continual support.

TO MY MOTHER, FATHER, AND SISTER

ABBREVIATIONS

aa amino acid

Ag antigen

AIDS acquired immune deficiency syndrome

ATP adenosine triphosphate

AZT 3' azidothymidine

BAL broncho-alveolar lavage

bp base pairs

CA capsid

CAT chloramphenicol acetyl transferase

CDC Center for Disease Control

CMV cytomegalovirus

CNS central nervous system

CSF cerebrospinal fluid

CTL cytotoxic T-lymphocytes

DMSO dimethyl sulfoxide

DNA deoxyribonucleic acid

DNase deoxyribonuclease

EDTA ethylenediaminetetraacetic acid

env envelope

gag group antigen

gp glycoprotein

HIV human immunodeficiency virus

HLA human leucocyte antigen

HSV herpes simplex virus

HTLV human T-cell leukaemia virus

IFN interferon

Ig immunoglobulin

IL interleukin

IVDU intravenous drug user

kb kilobase

KS Kaposi sarcoma

LN lymph node

LTR long terminal repeat

M matrix protein

MAb monoclonal antibody

mRNA messenger ribonucleic acid

N nucleocapsid

OD optical density

PAGE polyacrylamide gel electrophoresis

PBMC peripheral blood mononuclear cells

PCR polymerase chain reaction

RNA ribonucleic acid

RT reverse transcriptase

SDS sodium dodecyl sulphate

SIV simian immunodeficiency virus

TB tuberculosis

TBE tris-borate ethylene-diaminetetraacetic acid

TCR T-cell receptor

Taq *Thermus aquaticus*

TNF tumour necrosis factor

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

GENERAL INTRODUCTION

1.1 HUMAN IMMUNODEFICIENCY VIRUS

HIV was identified as the aetiological agent of AIDS just over a decade ago (Barre-Sinoussi et al., 1983); in this relatively short time an enormous quantity of information has been gathered about the nature of the virus and the pathogenesis of the disease, however, much still remains to be understood. AIDS in humans has been associated with two closely related viruses, HIV-1 and HIV-2. These viruses have similarly organised genomes, but show only 40 - 50% homology at the nucleic acid level. There is now considerable evidence for a simian origin of HIV (Sharp et al., 1994). Indeed, HIV-2 is more closely related to certain simian immunodeficiency viruses (SIVs), and some isolates of HIV-2 appear to be indistinguishable from isolates of SIV from sooty mangabeys (Gao et al., 1992). The evolutionary time point at which these cross-species transmissions occurred still remains to be determined.

HIV is a member of the lentivirus group of retroviruses; the mature virion has a diameter of about 120 nm, and comprises a cone shaped core containing two copies of the (+) RNA genome, reverse transcriptase (RT), the protease enzyme and group specific antigen (*gag*) proteins (p7, p9, p17). This is surrounded by a phospholipid envelope derived from the host cell membrane and containing viral glycoproteins (gp120/gp41) (Figure 1.1).

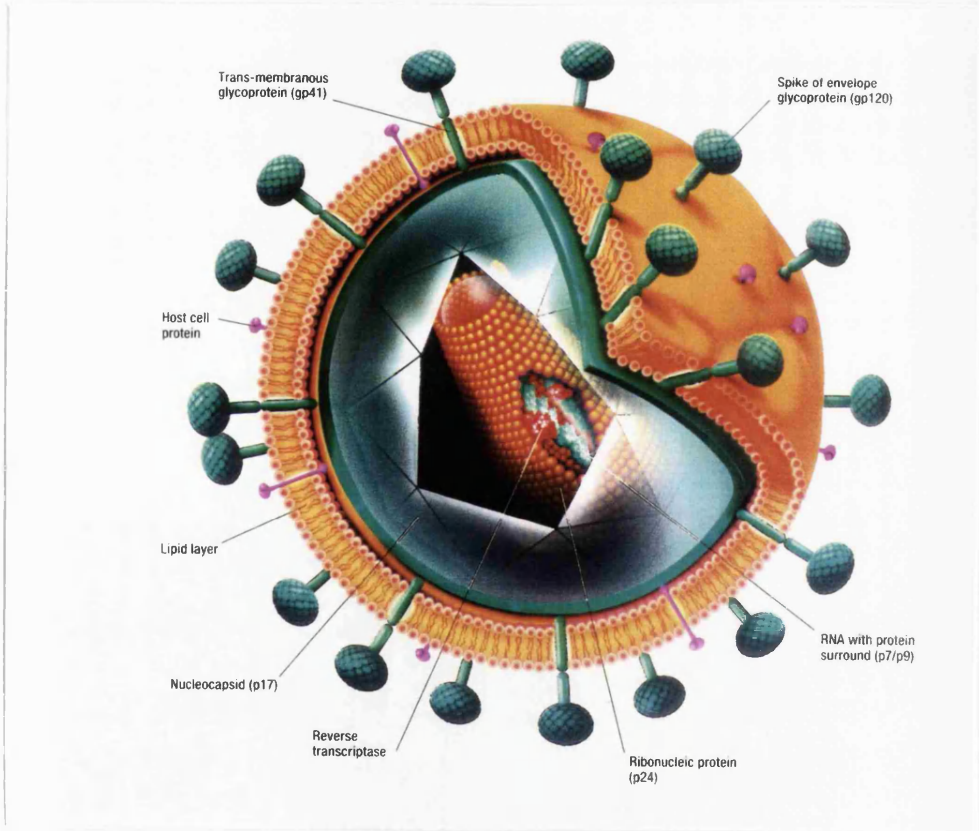


Figure 1.1: Schematic diagram of HIV-1. Reproduced from Boucher and Larder, 1994.

The genome of HIV-1 is 9.2 kb in length (Figure 1.2), and contains the three structural genes: *gag* (group specific antigen), *pol* (polymerase) and *env* (envelope glycoprotein) common to all retroviruses (Figure 1.2). In brief, the *gag* and *pol* genes are transcribed and translated to produce a large precursor protein, which is subsequently cleaved by the viral protease to produce the *gag* matrix protein (p17), the capsid protein (p24), and the nucleocapsid protein (p15) which is itself cleaved to produce the p7 and p9 proteins. Similarly, the following *pol* proteins are produced: protease (p10), reverse transcriptase (RT)/ribonuclease (RNase H) protein, and the p31 integrase/endonuclease protein. The *env* gene is transcribed and translated to produce the envelope glycoprotein precursor gp160, which is cleaved by a cellular protease to yield the surface glycoprotein, gp120, and the transmembrane protein, gp41. In addition, HIV-1 possesses two well-defined regulatory genes, *rev* (regulator of virion protein) and *tat* (trans-activator), and a third, *nef* (negative regulatory factor), whose role has not yet been fully established. The HIV-1 genome also encodes three genes thought to be involved in virion maturation and release, *vif* (virion infectivity factor), *vpr* and *vpu* (viral proteins R and U). A full description of HIV-1 replication will follow later in the introduction.

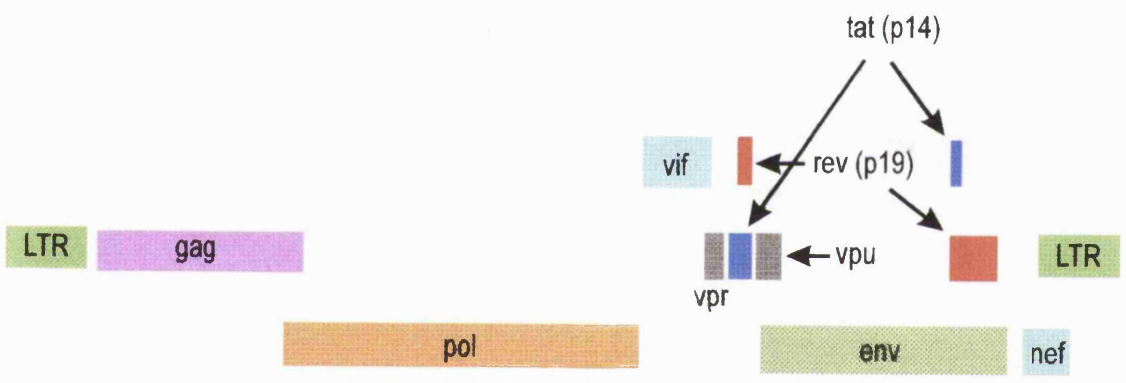


Figure 1.2: HIV-1 genome organisation.

1.2 HIV-1 REPLICATION

The HIV-1 life cycle consists of similar steps to those followed by all retroviruses:

1- Attachment of the virion to specific cell-surface receptor(s).

2- Penetration of the virion core into the cell.

3- Reverse transcription within the core structure to produce a double stranded DNA copy of the positive sense RNA genome.

4- Transport of the DNA associated with the virion proteins to the nucleus.

5- Integration of the viral DNA into the cellular DNA to form the provirus.

6- Synthesis of the viral RNA by cellular RNA polymerase II using the provirus as the template (transcription).

7- Processing of the transcripts into genomic RNA (full-length transcript) and messenger RNA (multiply spliced and incompletely spliced transcripts) for regulatory and structural proteins.

8- Synthesis of viral proteins.

9- Assembly and budding of virion.

10- Processing of capsid and envelope proteins.

A unique feature of the HIV-1 life cycle is the intricate regulation of virus expression by viral encoded proteins. The HTLVs, and Spumaviruses are the only other group of retroviruses that have similar but less elaborate controls.

In the following sections, the above feature will be highlighted while describing in more detail each step of the HIV-1 life cycle. Particular attention will be drawn to the transcription of the provirus which is finely controlled through the long terminal repeat (LTR) by both viral (*tat, nef*) and a broad range of cellular transcription factors.

1.2.1 Receptor-mediated activation of fusion

The HIV-1 life cycle begins with the binding of the viral envelope (gp120) to the cellular receptor known as CD4 (Dalgleish et al., 1984; Klatzmann et al., 1984; Maddon et al., 1986). The requirement for a specific receptor dictates the tropism of the virus and hence the major target cells which HIV infects are the T-helper lymphocytes (CD4+ lymphocytes) and the cells of the monocyte/macrophage lineage.

It is important to note that HIV-1 is polytropic since it also infects specialised cells such as the microglial cells present in the nervous tissues (Koenig et al., 1986; Wiley et al., 1986), the Kupffer cells of the liver, mucosal cells of the lining of the bowel and endoterium, as well as a large number of cells from the haematopoietic system, the skin, the brain, the bowel, and other tissues of the body (see Table 1.1).

The mechanism of viral entry has not been fully elucidated, however, it is clear that HIV-1 does not enter cells through the receptor-mediated endocytosis process since entry into cells is independent of acidification of endosomes (Stein et al., 1987; Maddon et al., 1988) and does not require the cytoplasmic domain of CD4 (Bedinger et al., 1988).

Table 1.1: Human cells suceptible to HIV *.

Hematopoietic:	**Bowel:**
T lymphocytes	Columnar and goblet cells
B lymphocytes	Enterocromaffin cells
Macrophages	Colon carcinoma cells
NK cells	
Megakaryocytes	**Other:**
Eosinophils	Myocardium
Dendritic cells	Renal tubular cells
Promyelocytes	Synovial membrane
Stem cells	Hepatocytes
Thymocytes	Hepatic sinusoid endothelium
Thymic epithelium	Kupffer cells
Follicular dendritic cells	Dental pulp fibroblasts
Bone marrow endothelial cells	Pulmonary fibroblasts
	Fetal adrenal cells
	Adrenal carcinoma cells
Brain:	Retina
Capillary endothelial cells	Cervix-derived epithelial cells
Astrocytes	Cervix (epithelium)
Macrophages (microglia)	Prostate
Oligodendrocytes	Testes
Choroid plexus	Osteosarcoma cells
Ganglia cells	Rhabdomyosarcoma cells
Neuroblastoma cells	Fetal chorionic villi
Glioma cell lines	Trophoblast cells
Neurons	
Skin:	
Langerhans cells	
Fibroblasts	

*:Susceptibility to HIV was determined by in vitro or in vivo studies (Levy, 1994).

The *env* gene of HIV-1 codes for a single chain glycoprotein precursor, gp 160, which is cleaved to yield the mature glycoproteins gp120 and gp41. These two subunits assemble into a heterodimer held together by noncovalent interactions, and are expressed at the virion surface in a functional, multimeric form. HIV-1 infection of CD4+ cells is initiated by binding of the virus to the cell surface via a high-affinity interaction between the first domain of CD4 and the HIV-1 outer envelope glycoprotein, gp120. CD4 consists of four extracellular immunoglobulin-like domains, a transmembrane segment and a intracytoplasmic tail. The region of CD4 directly implicated in interaction with gp120 is the V1 domain of the molecule (Figure 1.3), specifically, the first 106 amino acids (aa), for review see (Eiden and Lifson, 1992). Experiments using mutant CD4 molecules, substituting amino acids from murine CD4 that do not bind gp120 with human residues at homologous positions together with experiments using anti-CD4 monoclonal antibodies and site-directed mutagenesis , have implicated the region of CD4 homologous to the complementary determining region 2 (CDR2) of immunoglobulins in specific binding interactions with gp120. In other experiments, synthetic peptides derived from the CDR2-like domain of CD4 were shown to bind gp120, however peptides derived from the CDR3-like domain of CD4 blocked HIV infectivity, viral induced cell fusion, and gp120 binding indicating that the CDR3-like domain of CD4 also plays a critical role in gp120 binding. The use of soluble recombinant CD4 (sCD4) as a receptor mimetic has allowed the analysis of receptor binding (see above) and post-binding events which result in virus-cell membrane fusion.

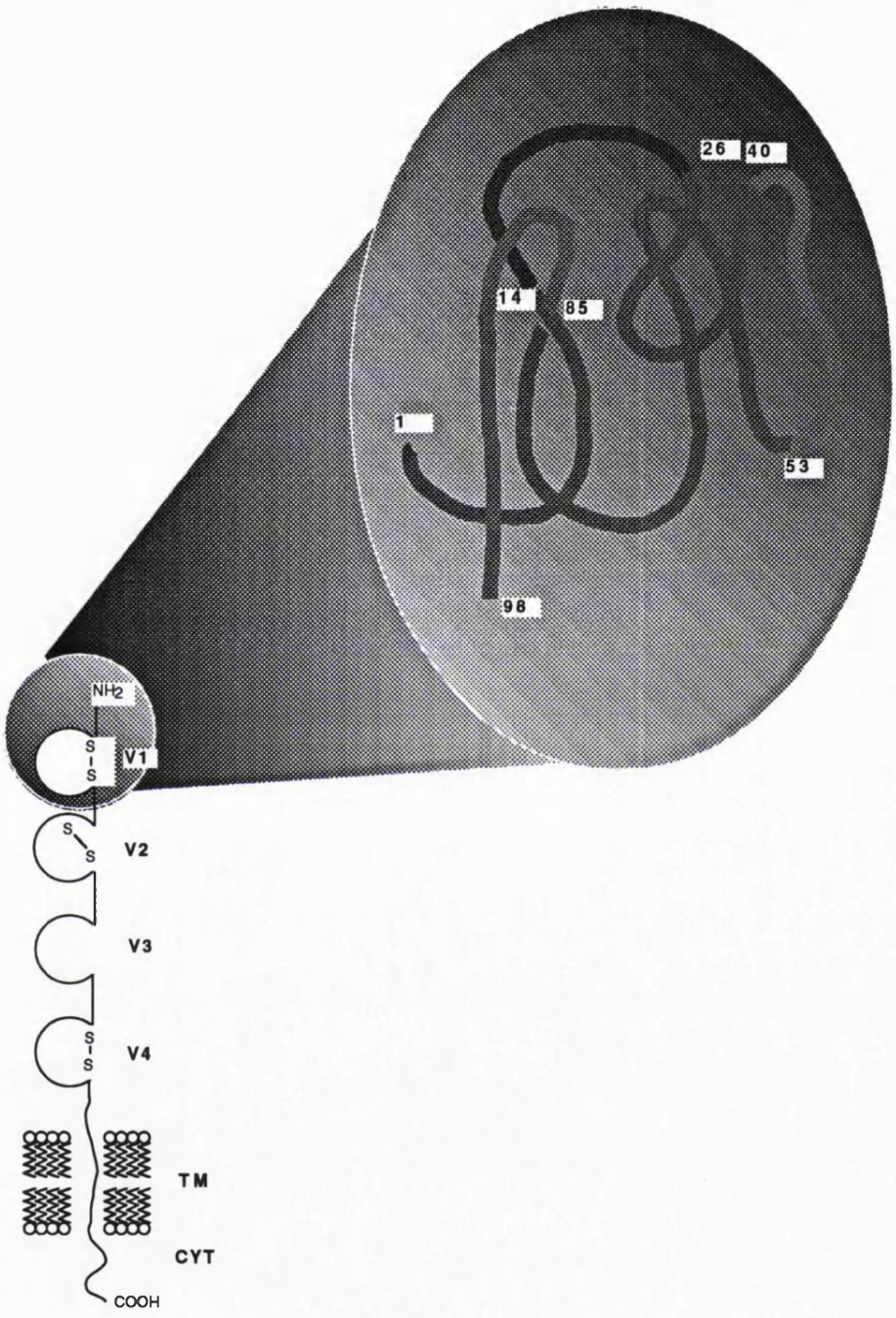


Figure 1.3: Schematic representation of the CD4 molecule. Detail of the V1 domain with amino acid regions 1 to 14, 14 to 26, 40 to 53, and 85 to 98 represented in different shades of grey (see text).

Soluble CD4 can specifically promote the dissociation of gp120 from gp41 in membrane-anchored, non-covalently associated gp120-gp41 complexes present either in virions or in the membranes of cells expressing HIV-1 envelope glycoproteins. This dissociation results in the exposure of fusogenic components of gp41. It is likely that *in vivo*, the exposure of this domain of gp41 at the interface between CD4 expressing cells and HIV-1 infected cells or free virions mediates virus-cell membrane coalescence, a process termed receptor-mediated activation of fusion. The use of monoclonal antibodies (MAbs) to gp120 and site-directed mutagenesis has led to the localisation of the CD4 binding site of gp120 to the carboxy terminal portion of the protein from residues 420 to 463. Numerous studies concentrating on HIV-1 envelope glycoprotein interactions with CD4 have produced discordant results, (for review see Signoret et al., 1993; and Eiden and Lifson, 1992). For example, different portions of the CD4 molecule as well as different regions of gp120 have been implicated in the binding and post-binding events and discrepancies between the interaction of gp120 with CD4 in solution and interactions between membrane-anchored CD4 and gp120-gp41 complexes are apparent. Finally, primary clinical HIV-1 virions cannot be effectively neutralised by soluble CD4 whereas laboratory adapted strains (T-cell line tropic) and T-cell tropic clinical isolates observed at the late stage of HIV-1 infection are neutralised.

The HIV-1 envelope V3 loop region has been identified as the major determinant of CD4 neutralisation sensitivity of HIV-1. Indeed, V3 loop sequences derived from primary isolates of HIV-1 were sufficient to confer on HTLV-IIIB, the degree of sCD4 neutralisation resistance observed for the HIV-1 isolates from which the V3 loop sequence had derived. The V3 loop sequence of primary isolates also conferred the macrotropic phenotype to the HTLV IIIB strain. These results demonstrated that the

tissue tropism and sCD4 neutralisation sensitivity of HIV-1 isolates are regulated by similar mechanisms and are dependent at least partly on the V3 loop sequence of gp120 (O'Brien et al., 1992; Hwang et al., 1992). The V3 loop of gp120 envelope protein is a disulfide bridged protein domain of about 35 amino acids that is the primary neutralisation determinant of HIV-1 as well as a major determinant for cell tropism. HIV-1 tropism will be discussed in more detail in sections 1.3.1 and 1.3.3. Mutations in the V3 loop or blockage by V3 MAbs have no effect on gp120 binding to CD4 but efficiently abrogate the post-binding events and hence the fusion of HIV-1 with susceptible CD4+ cells. Other studies have demonstrated the proteolytic cleavage of V3 under specific experimental conditions (Clements et al., 1991; Werner and Levy, 1993), however, *in vivo* cleavage of V3 subsequent to gp120 attachment to CD4 has not been demonstrated.

All these observations have been reconciled in a model for HIV entry (Signoret et al., 1993) in which the CD4-gp120-gp41 complex progresses through a series of conformational rearrangements in order to:

1- Induce the dissociation of gp120 from the gp120-gp41 complex;

2- correctly position gp41 relative to the surface membrane of the target cell (CD4 bending);

3- allow fusion of the virion envelope with the cell membrane.

The above requirements involve multiple progressive interactions of gp120 with CD4. These multivalent interactions would induce the necessary conformational changes of CD4 and gp120 molecules to allow fusion to occur.

An unanswered question in the mechanism of HIV-1 entry into its target cell is the identification of other cellular factor(s) which are required for fusion. For example,

mouse cells (NIH3T3) expressing the human CD4 gene will bind virus but are not able to mediate its entry and will not fuse with other cells expressing HIV-1 envelope proteins (Maddon et al., 1986; Ashorn et al., 1990). This defect can be overcome by forming hybrids with human cells (Dragic et al., 1992) indicating that a human component in the CD4+ cells is necessary for HIV-1 entry. The observation of proteolytic cleavage of gp120 in the V3 loop under particular experimental conditions has led to the hypothesis that a cellular protease may be the additional cellular factor required for HIV-1 entry. Callebaut and colleagues (1993), have proposed the cell surface protease CD26 as the human cell accessory factor (also referred to as co-receptor or effector but no consensus has been established) allowing HIV-1 to enter its target cells after binding. They showed that murine NIH 3T3 cells transiently transfected with CD4 and CD26 cDNAs were permissive to HIV-1 infection, and proposed that the human CD26, known as dipeptidyl peptidase IV (DPP IV), induces fusion after proteolytic cleavage of V3 (Callebaut et al., 1993). However, several groups have not been able to reproduce these results (Alizon and Dragic, 1994; Camerini et al., 1994; Patience et al., 1994; Broder et al., 1994; Lazaro et al., 1994). More investigations are required to elucidate the mechanism of entry of HIV-1 and identify all the cellular components necesary for virus penetration.

1.2.2 Reverse transcription

After entry of the core into the cytoplasm, the process of reverse transcription of the RNA genome into double-stranded DNA occurs, using the structural and enzymatic activities of the virion that entered the cell. Studies on the *in vitro* reverse transcription reaction have resulted in a detailed picture of the overall process. The viral genome has evolved to overcome two major problems in the process of reverse

transcription. Firstly, the need for RNA primers allowing precise end to end copying of the genome into DNA. Secondly, the requirement for the signals directing synthesis of RNA (viral genome) by RNA polymerase II to be located upstream of the initiation site of synthesis, outside the region to be copied. The above requirements are met by the presence of duplicated sequences at the 5' and 3' ends of the provirus forming a terminal redundancy known as the long terminal repeat (LTR). These LTR sequences contain *cis*-acting sequences recognised by the host cell transcription machinery and are necessary for integration and expression of the provirus.

The characteristic of reverse transcriptase (RT) allowing this metamorphosis is its ability, on encountering a block to DNA synthesis, to transfer the growing DNA chain to a similar sequence elsewhere, named "jumps" and then continue elongation. Two such jumps form the LTR by moving copies of the unique 5' and unique 3' regions flanking R, U5 and U3 respectively, to either end of the proviral DNA. The first transfer occurs after copying the short region consisting of R and U5 located upstream of the primer binding site (PB) forming the 5' end of the genome. These products, which are still attached to the tRNA primer at their 5' end, are called "strong stop" DNA, i.e., DNA synthesis starts by elongation from the 3' of the primer tRNA until the 5' end of the genome is reached. The strong stop DNA molecule then transfers to the 3' end of the genome where the other R sequence of the genome permits correct base pairing to allow DNA synthesis to continue. The jump is made possible by the RNase H activity of RT that removes the RNA from the RNA-DNA hybrid, leaving the minus-strand "strong stop" DNA free to base-pair with the 3' R sequence (Ben Artzi and Panet, 1996).

Once the jump has occured, the (-) strand DNA synthesis proceeds up to the 5' end of the template, which is the 5' end of PB because R and U5 have been removed by RNase H. The tRNA primer is still attached at the 5' end. The (-) strand DNA is also copied by RT. The primer molecule to initiate synthesis is created by a specific cleavage of the RNA template at the 5' end of U3 that leaves a fragment resistant to RNase H. The signal for this specific cleavage lies in the polypurine (PP) tract characteristic of all retroviruses. Initiation of (+) strand DNA synthesis has also been detected in other parts of the genome. These alternative initiation sites are mainly upstream sequences resembling the PP region, or pre-existing breaks in the genome. Incorrect initiation of the (+) strand DNA synthesis can result in the generation of viral DNA forms with either more or less than a full LTR. Such forms are frequently found in cloned viral molecules. The lack of a proper LTR will block integration, rendering such molecules over-represented among the unintegrated DNA molecules.

Initiation allows elongation of the (+) strand of DNA by reverse transcriptase up to the 5' end of the (-) strand. The (+) strand product is known as "positive strand strong-stop" DNA. Copying the first 18 bases of the primer-binding site generates a sequence that can form another template-primer hybrid with its complement at the 3' end of the (-) strand DNA. Synthesis of full-length, double-strand DNA can be completed after removal of the tRNA from the 5' end of the (-) strand. The genome RNA is completly degraded by the action of RNase H to allow synthesis of the (+) strand DNA (Figure 1.4). The sites and structure for DNA synthesis are not fully elucidated, however, it is known to occur in a defined structure including the capsid proteins, RT and integrase.

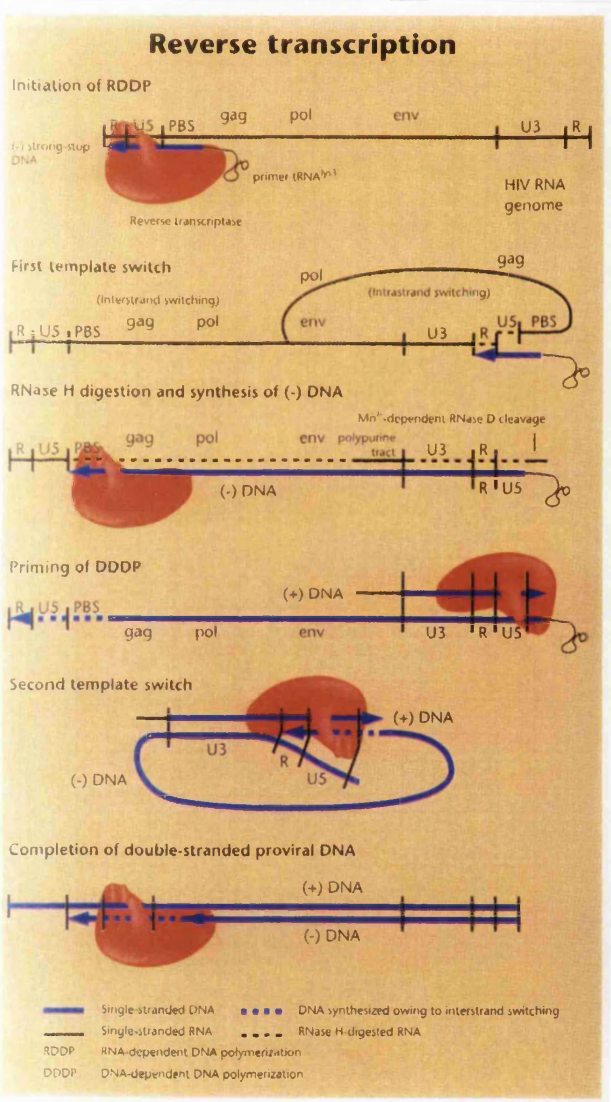


Figure 1.4: Reverse transcription process reproduced from Wainberg and Gu, 1995.

The structures of RT and RNAase H have been elucidated complexed with the non-nucleoside reverse transcriptase inhibitor (NNRTI), Nevirapin, and revealed an asymmetric dimer. The polymerase domain of the 66-kilodalton subunit has a large cleft analogous to the Klenow fragment of Escherichia *coli* DNA polymerase I (Figure 1.5). HIV-1 RT heterodimer shows a deep cleft that runs between the RNase H active site and the pol active site(Kohlstaedt et al., 1992). The crystal structure of RT has shed light on the mechanisms of drugs activity and occurrence of resistant mutations (Boucher and Larder, 1994).

Following DNA synthesis, the viral DNA is transported to the nucleus. Two additional forms of viral DNA have been identified: covalently closed circles containing either one or two LTR. The circles containing one LTR could be formed from linear molecules by homologous recombination across the LTRs, but it is more probable that they represent an aberration during DNA synthesis in which the (+) strand is not displaced from the (-) strand at the final step. The two LTR circles include a variety of structures arising either by ligation of the ends of the linear molecule or by integration of the viral DNA into itself.

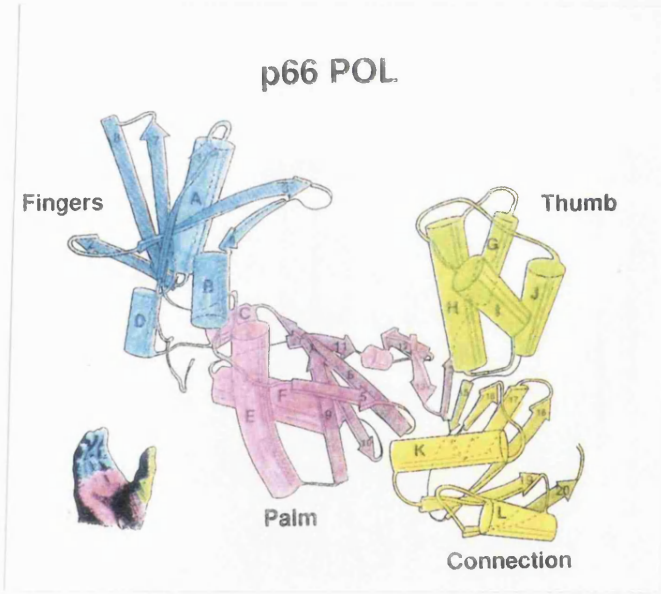


Figure 1.5: 3D structure of RT reproduced from Kohlstaedt et al., 1992. The polymerase domain of HIV-1 RT p66 is shown with helical regions represented as cylinders and ∃ strands as arrows. The different coloured region of the p66 are labelled above.

1.2.3 Mechanism of integration

Like reverse transcription, integration of viral DNA occurs in the context of a specific structure derived from the viral capsid. Integration of the double-stranded DNA copy of the retrovirus genome into a chromosome of the host cell requires the HIV integrase and specific DNA sequences at the ends of the viral LTR. The *in vitro* mechanism of integration has been elucidated by the isolation and characterisation of DNA intermediates obtained from infected cells while integration is still ongoing. The integration reaction proceeds via the following three steps: (i) 3' processing, in which integrase removes 2 nucleotides from the 3' end of each strand of linear viral DNA so that the viral 3' ends terminate with the CA dinucleotide; (ii) strand transfer, a coupled cleavage-ligation reaction during which integrase makes a staggered cut in the target DNA and ligates the recessed 3' end of the viral DNA to the 5' ends of the target cellular DNA at the cleavage site, producing a gapped intermediate and (iii) gap repair, in which the integration process is completed by removal of the two unpaired nucleotides at the 5' end of the cellular DNA and repair of the gaps between the viral and target DNA sequences, thereby generating the short direct repeats that flank the provirus. This gap repair of the initial staggered cut probably involves the cellular DNA repair system and generates the characteristic duplication of cell DNA flanking the provirus (Figure 1.6). The protein factors involved in the last step of the reaction remain to be characterised, whereas the first two steps of integration are integrase dependent with each step proceeding by a one-step transesterification mechanism (Engelman et al., 1991).

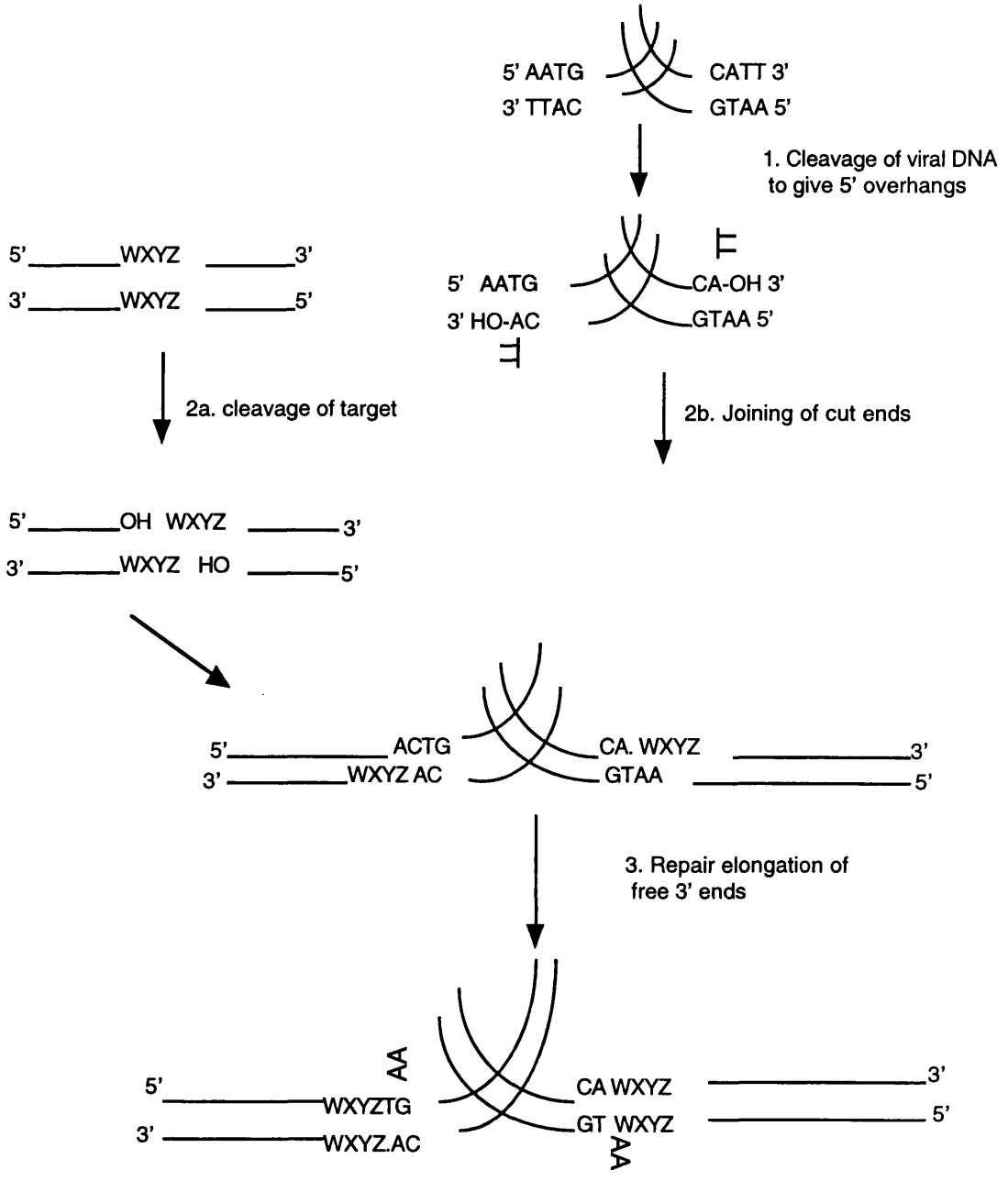


Figure 1.6: Mechanism of HIV-1 genome integration. The multiple step event has been subdivided into four steps which are labelled on the figure.

Once integrated, the viral DNA is referred to as the provirus. All further replication, namely, expression of the provirus (transcription/translation) is carried out using the cellular enzymatic machinery. Like other retroviruses, replication of HIV-1 is partly determined by the state of proliferation and activation of the infected cell (Gaynor, 1992; Levy, 1994).

Upon stimulation of the latently infected cell, synthesis of RNA occurs; these RNA molecules serve both as templates for viral proteins and as RNA genomes to be packaged into progeny virions (see later). Assembly of the immature virion occurs at the inner surface of the cell membrane, where the structural core proteins have aggregated. Simultaneous with the assembly of the capsid precursor is the accumulation of viral envelope proteins on the outer surface of the cell. The association of two viral genomic RNA molecules with the smaller core proteins also occurs, and the nucleoprotein complex is packaged into the particle at the plasma membrane. Once assembly is complete, the virus is released from the cell surface, where the envelope proteins undergo proteolytic cleavage (gp160≡gp120/gp41) by the cellular protease enzyme, forming the mature virion.

1.2.4 Control of transcription

The transcription of the provirus is orchestrated by viral and cellular factors through DNA regulatory sequences found within the HIV-1 LTR. The LTR comprises three regions: U3, R and U5; denoting sequences present in genomic RNA uniquely at the 3'end, repeated in the RNA at either end, or uniquely at the 5' end of viral RNA respectively. Viral transcription from integrated proviral DNA is carried out by cellular RNA polymerase II, which initiates in the upstream LTR at the U3-R border and terminates in the downstream LTR at the R-U5 border, yielding an RNA identical to

genomic RNA. This single RNA precursor is subsequently processed by (a) polyadenylation at the 3' end of R to yield a genome length molecule and (b) splicing (multiple or single) of a fraction of the transcripts to generate subgenomic mRNA species. The post-transcriptional regulation involving the viral protein *rev* will be discussed in section 1.2.5.

The HIV-1 LTR (Figure 1.7) consists of a 453 base pair U3 region, a 98 base pair R region and an 83 base pair U5 region. It contains characteristic regulatory elements: a polyadenylation signal sequence 19 bp from the R-U5 junction, the sequence ATATAAG or TATA box, 24 bp 5' of the cap site and a transfer RNA binding site complementary to the 3' end of the $tRNA^{Lys}$ located 3' of the 5' LTR which is used as a primer for (-) DNA strand synthesis. Within the U3 and R regions are multiple cis-acting elements involved in HIV-1 gene expression, hence, the LTR functions as a promoter-enhancer of HIV-1 transcription (Gaynor, 1992). The LTR can be broadly divided into, modulatory, core promoter and transactivating regions. The modulatory region include *cis*-acting elements with homology to the AP-1 enhancer motif between positions -347 and -329 relative to the cap site, two NF-AT sites (Nuclear Factor of Activated T-cells) between positions -292 and -251, an Upstream Stimulatory Factor (USF) site positioned between -139 and -124 and two NF-kB enhancer elements positioned between -104 and-81. The core region is composed of three Sp1 binding sites located between positions -78 to -47 and the TATA box (-28 to-24). The region of the LTR from +1 to +59 confers transcriptional transactivating activity by the viral protein *tat*. This transactivating region, namely *Tat* Response element (TAR), folds into a stable RNA stem-loop structure in the 5' termini of viral transcripts (Figure 1.8). Other regulatory sequences that govern HIV-1 replication lie immediately downstream of the LTR and include the primer binding

site, major 5' slice donor, and genomic RNA packaging signal (P) (Gaynor, 1992). All the different transactivating protein binding sites are critical in altering HIV-1 gene expression in response to changes in cellular signal transduction pathways, affecting both basal and *tat*-induced gene expression. Hence, HIV-1 is subject to many of the same regulatory signals that are important in controlling cellular gene expression. The regulation of HIV transcription, like that of cellular genes, represents the result of multiple interactions of *trans*-activators and -repressors with the basal transcriptional apparatus.

1.2.5 Cellular transcription factors and their specific binding sites

The majority of data concerning the activity of the HIV-1 LTR has been derived from the study of laboratory-adapted molecular clones or molecular deletion mutants (Gaynor, 1992). Studies on naturally occurring LTR genotypes have received less attention. The studies on LTR genotype/phenotype analysis will be reviewed in a subsequent section on HIV-1 variability (section 1.3.4). It is important to note that less dramatic effects are seen on viral growth kinetics with enhancer mutations in whole virus molecular clones compared with effects seen using transfections of HIV-1 reporter constructs containing mutations in the enhancer region. It is therefore important to scrutinise *in vitro* functional data and be cautious when extrapolating these observations to possible *in vivo* effects.

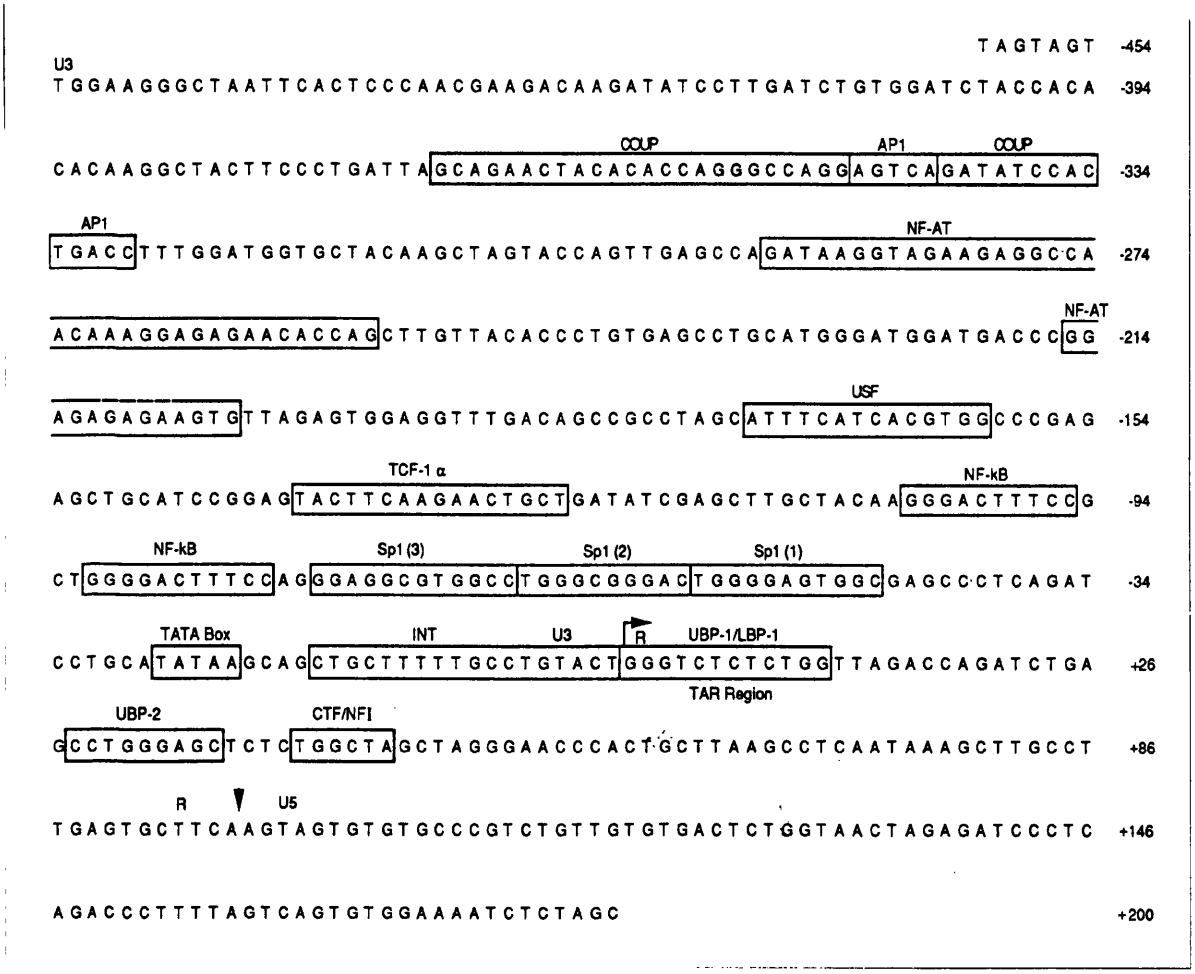


Figure 1.7: HIV-1 LTR nucleotide sequence and the characterised transcription factor binding sites. All transcription factors are discussed in the text. Reproduced from Gaynor 1992.

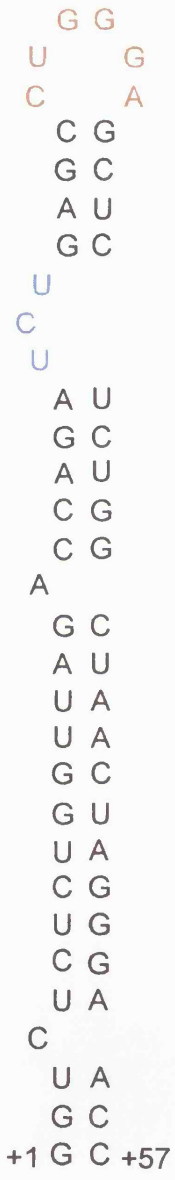


Figure 1.8: Secondary structure of the TAR element. The TAR buldge is shown in blue, and the TAR loop is shown in red.

Transcription factors do not function independently of each other (i.e. in isolation), on the contrary, they form regulatory networks in which several factors interact to regulate gene transcription.

1.2.5.1 AP-1: Activator Protein 1

AP-1 is a homodimer of *Jun* or a heterodimer between members of the *Jun* and *Fos* families. Both *c-jun* and *c-fos* are nuclear oncogenes and transcription factors which contain a structural feature known as a leucine zipper, which allows dimerisation between the jun and fos proteins. AP-1 is involved in the expression of a variety of cellular and viral genes via its interactions with the consensus DNA recognition motif "TGACTCA" found in cellular and viral promoter elements. The *jun* and *fos* proteins can interact with other cellular proteins to modulate gene expression, and can form complexes with steroid and steroid-related hormone receptors resulting in the inhibition of hormone receptor mediated transcriptional activation. Thus, AP-1 is capable of both positive and negative effects on gene expression. The AP-1 binding sites in the HIV-1 LTR are found within the binding sites for members of the Chicken Ovalbumin Upstream Promoter (COUP) family of transcription factors which are members of the steroid/thyroid hormone receptor superfamily. AP-1 could inhibit COUP binding to the HIV-1 LTR, and therefore its *trans*-activation capacity in two ways: it could directly interact with COUP factors or compete for binding to overlapping regulatory motifs. The precise role of AP-1 in the regulation of HIV-1 gene expression has yet to be established. It is important to note that the HIV-1 LTR is under the influence of an array of factors that *in vivo* could compete with each other.

1.2.5.2 COUP-TF: Chicken Ovalbumin Upstream Promoter-Transcription Factors

COUP-TF belong to the steroid/thyroid hormone receptor superfamily. They interact with a non-DNA binding cellular factor (S300II) to activate *in vitro* transcription of the ovalbumin promoter. The HIV-1 LTR contains a palindromic recognition sequence, which binds the 68kDa form of COUP-TF. Unlike the case of the ovalbumin promoter, mutations in the COUP-TF binding sites in the HIV-1 LTR results in a 2 to 3 fold increase in HIV-1 gene expression, which is consistent with the potential negative regulatory role of this element (Orchard et al., 1990; Cooney et al., 1991). Several variant HIV-1 strains have been identified that contain altered sequences in the HIV-1 COUP element and affect the binding properties of COUP-TF in both a positive and negative manner (Cooney et al., 1991), and, (Ait-Khaled et al., unpublished observations, see chapter 5).

1.2.5.3 NF-AT: Nuclear Factor of Activated T cells

Two purine-rich sequences in the HIV-1 LTR from -253 to -213 relative to the start of transcription serve as binding sites for a cellular transcription factor known as NF-AT. NF-AT is present at low levels in unstimulated T-cells, but its binding is markedly increased upon activation of T-lymphocytes. Other cellular factors also bind to these purine-rich sequences and are potentially capable of modulating HIV-1 gene expression. The role of the NF-AT sites in regulating HIV-1 gene expression upon activation of T-lymphocytes has been demonstrated using transfection studies with reporter constructs containing deletion or site-directed mutations of the NF-AT site (Lu et al., 1990). The results showed that the sequences recognizing the NFAT-1 and USF cellular transcription factors contribute to the negative regulatory effect of the HIV LTR NRE. NF-AT also activates interleukin-2 (IL-2) gene expression through a specific interaction in the IL-2 gene promoter. NF-AT is composed of two

subunits: a constitutive nuclear factor and T-cell specific cytoplasmic factor. The immunosuppressive agents cyclosporin and FK506, which associate with cytoplasmic receptors (cyclophilin and FKBP, respectively), inhibits DNA binding by NF-AT, possibly by preventing the interaction between the T-cell cytoplasmic and nuclear components. These inhibitors act by inhibiting the activity of the calcium/calmodulin-dependent protein phosphatase calcineurin. The gene encoding the cytoplasmic subunit of NF-AT (NF-ATc) has recently been cloned (Northrop et al., 1994). Homology has been observed between NF-ATc and Rel/NF-kB family members. The nuclear subunit of NF-AT (NF-ATn) is believed to contain products of the *jun/fos* family, which normally make up the AP-1 transcription factor. Full activation of NF-AT results from a phorbol myristyl acetate (PMA)-dependent mechanism involving synthesis of the components of the AP-1 transcription factor, and a calcium-dependent mechanism involving nuclear translocation of NF-ATc. However, inhibitory effects mediated by the Negative Regulatory Element (NRE) cannot be ascribed to the action of a single factor and are probably cell lineage dependent.

1.2.5.4 USF: Upstream Stimulatory Factor

USF binds to the sequence "CACGTG", known as the upstream stimulatory core sequence. USF is a 43 kDa nuclear protein capable of dimerisation and so could potentially interact with other cellular factors. This element is part of a binding motif in the HIV-1 LTR that binds negative regulatory proteins. Gel retardation and UV crosslinking experiments have demonstrated the binding of a cellular protein to a region of the LTR extending from -159 to -173 upstream of the transcription start site which has a USF recogniton sequence.

HIV-1 constructs with a deletion in the LTR that includes the USF element show increases in both gene expression and growth kinetics (Lu et al., 1990). Thus, the USF binding site may serve as a negative element within the HIV-1 LTR. However, in the presence of cytomegalovirus (CMV) immediate-early proteins, this element is required for synergistic activation of the HIV-1 LTR (Ghazal et al., 1991). Hence, depending on which cellular factors interact with this regulatory element, both positive and negative effects could occur on HIV-1 gene expression. The binding ability of USF is abolished by the mutation of the core recognition sequence. A recent study showed that upon binding the LTR undergoes a structural distortion. The ability of USF to bend the LTR template upon binding was demonstrated by the circular permutation assay, a method based on the position-dependent effect of DNA bends on the electrophoretic mobilities of DNA fragments. Upon addition of USF, transcription from constructs containing an intact binding site was increased, while responsivness in constructs with a mutated sequence was abolished. Addition of a decoy plasmid which contains multiple repeats of the USF binding sequence down-regulated the transcription from the LTR. These results suggest that USF is a positive regulator of LTR-mediated transcriptional activation (Di Fagagna et al., 1995). The *in vivo* function of the USF site could necessitate the presence of a complex combination of other DNA-binding factors, a situation which is not reproducible in the *in vitro* transcription assays. More experiments are needed to clarify these points.

1.2.5.5 TCF-1 α: T-cell factor-1 α

The region of the LTR between positions -123 and -138 binds a 55kDa cellular transcription factor known as TCF-1 α. This T-cell specific cellular transcription factor also binds to pyrimidine-rich sequences in the T-cell receptor α gene.

Functional studies of the TCF-1 α gene product in transfection experiments indicate that it activates gene expression by interacting with its binding site in the T-cell receptor promoter. No data are at present available regarding the role of this factor on HIV-1 gene expression. However, primary isolates with a duplicated TCF-1 alpha site have been shown to have increased growth properties in T-lymphocytes (Golub et al., 1990), and HIV proviral DNA obtained directly from infected individuals has also been shown to contain TCF-1 α duplications (Koken et al., 1992), (see Chapter 5). Thus, the TCF-1 alpha regulatory element may be important in regulating HIV-1 growth properties *in vivo*.

1.2.5.6 NF-kB: Nuclear factor Kappa B

The NF-kB elements are the most widely studied sites in the enhancer region of the HIV-1 LTR. The two NF-kB motifs are important for HIV-1 promoter activity both in activated T lymphocytes and other cell lines. NF-kB proteins are present constitutively in non B-cells bound to a cytoplasmic inhibitory protein known as Ik-B. NF-kB is normally associated with its inhibitor IkB. However, phosphorylation of IkB by protein kinase C (PKC) dissociates IkB from NF-kB and allows its translocation to the nucleus, where it binds the LTR of the HIV-1 provirus amongst other promoter elements. NF-kB comprises a collection of related proteins of different sizes, including p50, p50B, p65, and p85(c-rel). The rel homology region of these proteins contains sequences important for DNA binding, formation of the various homo- and heterodimers, and nuclear localizaton. In the case of the HIV-LTR, the active NF-kB is a heterodimer composed of p50 and p65 subunits, where p65 is responsible for the transcriptional activity of the complex; p65 represents the major rel/kB transcriptional factor detected in the nuclei of the T lymphocytes upon activation. Resting T cells do not contain nuclear NfkB. Extracellular signals such as TNF, IL-1,

T-cell activation, or phorbol esters (such as PMA) act through signal transduction pathways, including ceramide, reactive oxygen intermediates, and protein kinases

Using LTR-reporter gene plasmids, several studies have demonstrated that a large number of stimuli are capable of inducing HIV transcription through NF-kB (Nabel et al., 1988; Griffin et al., 1989; Folks et al., 1989). A direct effect of NF-kB on the induction of viral production has been observed in the context of cell lines chronically infected with HIV. Treatment of these cells with phorbol esters or TNF-alpha results in induction of NF-kB binding to the LTR and augmented HIV gene expression (Folks et al., 1989). Thus, the LTR NF-kB binding sites play an important role in the activation of latent proviruses. Their role in acute productive HIV replication was thought to be far less important since infectious HIV DNA clones containing mutations or deletions of the LTR NF-kB sites replicated as efficiently as wild-type proviruses in PHA-activated lymphocytes (Leonard et al., 1989). Until recently, the majority of studies analyzing mechanisms of viral transcription had been performed in lymphoblastoid T- cell lines which are not necessarily be the best model for HIV replication *in vivo*. An elegant study by Alcami and collaborators has compared the requirements for k-B dependent activation of the HIV LTR in peripheral blood CD4 T lymphocytes and lymphoblastoid T cells using transient transfections. In addition, these workers analysed whether NF-kB dependent LTR transactivation was obligatory for both initiation and maintenance of viral transcription in blood CD4 T lymphocytes infected with either kB-mutated or wild-type HIV proviruses (Alcami et al., 1995). The results showed that the basal activity of the HIV LTR in unstimulated CD4 T lymphocytes was extremely low, i.e., 81-fold less than that observed in the lymphoblastoid cell line JJHan. The very low levels of spontaneous transcription observed with mutated NF-kB LTR vectors in resting

CD4 blood lymphocytes demonstrated that basal transcription is NF-kB dependent. In both primary CD4 T lymphocytes and CD4 T-cell lines, a significant increase in LTR activity was induced by phorbol myristate acetate (PMA). In both cell types this increase was mediated by induction of NF-kB since deletion of NF-κB binding sites abrogated the increase. The results also showed that in contrast to the situation in the lymphoblastoid cell lines, in resting blood T lymphocytes, deletion of NF-kb responsive elements from the LTR completely prevented *tat*-induced transactivation. This confirmed previous observations suggesting an indirect interaction of *tat* with the viral enhancer (Berkhout et al., 1990; Harrich et al., 1990; Taylor et al., 1992) which may explain the synergistic effect of *tat* and the Nf-kB complexes on transcription initiation of the HIV LTR. The NF-kB-dependent *tat* transactivation of the LTR in resting lymphocytes is in agreement with data by Moses and colleagues obtained in unstimulated human primary monocytes/macrophages co-transfected with *tat* and HIV LTR-CAT vectors with critical point mutations in the viral enhancer (Moses et al., 1994). These results highlight the importance of using non-transformed cells for the analysis of HIV transcription and replication. As opposed to T lymphoblastoid cells, the environment of normal CD4 human lymphocytes, the natural target of the virus, does not support any constitutive activity of the LTR and is not permissive to spontaneous HIV-1 genome transcription. Both these studies are in good accord with the long-term quiescence observed in the resting T lymphocytes and monocytes of the infected patients.

1.2.5.7 Sp1

Three binding sites for the mammalian transcription factor, Sp1, have been identified 5' to the HIV TATA box (Jones et al., 1986). These DNA binding sites were demonstrated by DNA footprinting using the 95- to 105-kD Sp1 protein purified from

HeLa cell nuclei, and corresponded to a GC rich region of the LTR spanning positions -46 to -78bp. Sp1 is a strong transcriptional activator that contains three zinc finger motifs involved in DNA binding and two glutamine-rich transcription activation domains. The three Sp1 sites in the HIV LTR are labelled sites I, II, and III from the 3'end. Transient transfection assays using HIV LTR-CAT plasmids have shown that mutations or deletion of the Sp1 sites results in reduced basal transcription as well as reduced *tat* transactivation (Harrich et al., 1990). The initial binding studies showed that Sp1 site I (the most 3' site) binds to purified Sp1 with lower affinity (Jones et al., 1986). This was confirmed by functional studies which demonstrated that HIV deleted in all three Sp1 sites replicated with similar kinetics to virus containing inactivating point mutations in sites II and III (Leonard et al., 1989; Parrott et al., 1991). The Sp1 binding sites have also been shown to be involved in *tat*-mediated transactivation (Berkhout et al., 1990; Harrich et al., 1990) and physical interaction between *tat* and Sp1 has been reported (Jeang et al., 1993). More recently, the strict requirement for the Sp1-DNA binding sites for both basal and *tat*-mediated transactivation of the HIV LTR has been shown using an *in vitro* cell-free transcription system (Sune and Garcia-Blanco, 1995). Nevertheless, as discussed for NF-kB, these data have to be analysed with caution since in the natural environment of the provirus the Sp1 sites could be dispensible.

1.2.5.8 TATA box and Leader Binding Protein-1 (LBP-1)

The HIV LTR like most mammalian class II RNA polymerase promoters contains a TATA box which is located 24 bp 5' to the start site of RNA transcription. Specific transcription initiation at class II gene promoters requires the assembly of a precisely positioned RNA polymerase II (RNA Pol II)-containing preinitiation complex (PIC) over the core promoter DNA region. The process of PIC formation on TATA box

involves the ordered assembly of at least seven basal transcription initiation factors including TFIIA, -B, -D, -E, -F, -G/J and -H, in addition to RNA Pol II. TFIID is the only factor with sequence-specific DNA-binding activity and initiates PIC assembly through recognition of the TATA box, reviewed by (Zawel and Reinberg, 1992). Native TFIID is a large multisubunit complex composed of a small TATA-binding protein (TBP) and several tightly associated factors (TAFs), reviewed by (Chiang et al., 1993). The process of transcription by RNA polymerase II begins with TATA binding protein (TBP)- mediated recognition of a TATA element by a general initiation factor, transcription factor IID (TFIID), forming a multiprotein-DNA complex that coordinates assembly of class II initiation factors and Pol II into a PIC. TFIIB is the second general transcription factor to enter the PIC, creating a TFIIB-TFIID(TBP)-DNA platform which is in turn recognised by a complex of Pol II and TFIIF (pol/F). Once PIC assembly is complete and in the presence of nucleoside triphosphates, strand separation at the transcription site occurs to give an open complex, the carboxy-terminal domain of the large subunit of Pol II is phosphorylated, and Pol II initiates transcription and is released from the promoter. During elongation, TFIID can remain bound to the core promoter supporting rapid reinitiation of transcription by Pol II and the other factors (Zawel et al., 1995). Core-promoter binding by the TBP subunit of TFIID is an intrinsically slow step because of the marked DNA deformation induced in the TATA element. A recent study reported the crystal structure of a TFIIB-TBP-TATA-element ternary complex providing the first structure of a protein recognizing the TBP-DNA complex (Nikolov et al., 1992; Nikolov and Burley, 1994). The strcture of this complex is in good accord with some aspects of the known biochemical and transcriptional properties of the PIC.

Mutation of the LTR TATA box reduces, but does not eliminate, basal transcription of HIV RNA in transient transfection assays (Berkhout and Jeang, 1992; Nabel, 1988). TATA box mutations in HIV proviral DNA result in loss of infectivity (Lu et al., 1989). This apparent discrepancy may be explained by the role of the TATA box on the *tat*-mediated transactivation of the LTR. Indeed, *tat*-transactivation is required for efficient HIV gene expression typical of productive replication (see below, section 1.2.5.9), and the nature of the basal transcription complex formed at the TATA box would determine the ability of *tat* to activate LTR transcription. In addition to providing the substrate for interaction with *tat*, the PIC also interacts with cellular factors binding upstream of the TATA box, such as Sp1 and NF-kB (Li et al., 1991). The LTR region spanning positions -16 to +27 bp which comprises the TATA box, also binds LBP-1 (leader binding protein 1) (Jones et al., 1988)or UBP-1 (upstream binding protein) (Wu et al., 1988). LBP-1 has been reported to induce both positive and negative effects on transcription, however, the *in vivo* role of LBP-1 remains unclear.

1.2.5.9 TAR element and interactions with *tat*

The viral gene product of HIV, *tat*, is essential for HIV replication, and transactivates gene expression through a responsive element called TAR. The TAR element was originally mapped to the LTR between nucleotides -17 and +80 of the LTR and thus includes sequences present at the 5' end of all HIV mRNAs (Rosen et al., 1985b). The mRNA transcribed from the TAR region can form a stable stem-loop structure which has been shown to be the functional response element of *tat* (Figure 1.8), it is located at nucleotide positions +1 to +57 (Berkhout et al., 1989). One of the primary effects of *tat* is to facilitate elongation of initiated transcripts which would otherwise terminate shortly beyond the 3' end of TAR. These prematurely

terminated transcripts 55 to 70 bases long, accumulate in uninduced cells and reflect non-productive initiation events from the HIV promoter. Mutations disrupting the sequence and structure of the TAR RNA stem inhibit the generation of short transcripts. The production of short transcripts is independent of theTAR loop, but highly dependent of the TAR RNA stem (Kao et al., 1987; Laspia et al., 1989; Marciniak et al., 1990a; Ratnasabapathy et al., 1990).

Although *tat* binds to the pyrimidine bulge (the 3-nucleotide bulge of TAR RNA ,+23 to +25(UCU)) and the immediate flanking base-paired sequences of TAR RNA *in vitro* (Figure 1.8), this interaction is not sufficient to activate transcription *in vivo* (Nabel, 1988; Dingwall et al., 1990; Roy et al., 1990). This suggested that host cell factors were involved in *tat* transactivation through TAR. Indeed, one study has demonstrated that levels of HIV-1 transcriptional activity vary up to a 1,000-fold between cell types (Barry et al., 1991). Moreover, studies using rodent-human cell hybrids showed that whereas certain rodent cells, such as CHO cells either do not support *tat*-mediated transactivation or do so very poorly, *tat* transactivation is restored in rodent-human hybrid cells that retain the human chromosome 12 (HuChr12) (Hart et al., 1989; Newstein et al., 1990). The native TAR RNA loop sequences (+31 to +34) are required for HuChr12 to support transactivation (Hart et al., 1993), and recently, the same investigators have shown that either a species-specific 83kDa TAR RNA loop-binding protein is directly encoded on HuChr12 or a HuChr12 protein(s) induces the expression of a 83-KDa TAR-binding protein in nonprimate cells which is responsible for the HuChr12 dependent *tat* transactivation (Hart et al., 1995).

Other investigators have identified a 68-kDa HeLa cell nuclear protein which binds to the loop of TAR and increases transactivation (Marciniak et al., 1990a; Marciniak et al., 1990b). Using RNase protection mobility shift assays and UV cross-linking, other workers have shown that a 140-kDa HeLa nuclear protein binds specifically to TAR RNA; by mutational analysis of TAR RNA, they showed that the sequence and structure of stem regions I (bases 1 to 9), II (bases 10 to 15), and III (bases 17 to 21), the pyrimidine bulge, and the non-base paired adenine (position +16) are important for the above interaction, whereas neither the loop, nor the stem region IV are essential for the binding of this host cell factor to TAR RNA (Rounseville and Kumar, 1992). The above protein was referred to as stem binding protein or SBP. Recently, it has been shown that TAR RNA/p-140 SBP binding activity was low in resting unstimulated human T-lymphocytes but increased within 2 h of mitogen-activation and remained high for at least 48 h (Rothblum et al., 1995). Using similar techniques, another study has reported two HeLa cell nuclear proteins (TRP-185 and TRP-140) which bind to TAR RNA. TRP-185 binding is dependent on the loop sequences of TAR and requires an additional host cell factor, whereas TRP-140 is apparently nonspecific (Wu et al., 1991). Although several cellular factors have been shown to interact with TAR RNA and/or *tat* their role in transactivation remains unclear. More recently, it has been demonstrated that *tat* binds directly and specifically to the basal transcription factor TFIID *in vitro*, the complex is formed by *tat* and holo-TFIID (TBP and associated factors (TAFs)) (Kashanchi et al., 1994). *Tat* binds, through aa 36-50, directly to the TBP subunit of TFIID. The above finding suggests that one mechanism of *tat* transactivation may be in stabilizing the initiation complex, the way *tat* increases its processivity is unclear, but it may facilitate the assembly of a more processive elongation complex (Kashanchi et al., 1994). Using an *in vitro* kinase assay, a cellular protein kinase activity has been

identified, this *tat*-associated kinase (TAK) specifically binds to the activation domains of *tat* proteins (Herrmann and Rice, 1993). The afore mentioned investigators have now identified an *in vitro* substrate of TAK, namely, the carboxyl-terminal domain of the large subunit of RNA pol II (Herrmann and Rice, 1995). They also showed that TAK activity is inhibited by dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB), a nucleoside analogue that has been shown to selectively inhibit *tat* function *in vitro* and *in vivo*. Phosphorylation of the carboxyl-terminal domain has been proposed to trigger the transition from complex assembly to active elongation and to influence events during the elongation phase of transcription. From these results, the authors propose that TAK is an important cellular co-factor that mediates *tat* function (Herrmann and Rice, 1995). Finally, the cloning and characterization of a cellular protein that interacts with HIV-1 *tat* has been reported. This protein named TAP (*tat*-associated protein), is a highly acidic 209-amino-acid protein, and deletion analyses have identified a TAP C-terminal region rich in acidic amino acids and leucine residues which acts as a strong transcriptional activator in heterologous expression assays; thus, TAP binds *tat in vitro* and *in vivo* (Yu et al., 1995b). The same group has also shown that TAP binds the general transcription factor TFIIB, and that the C-terminal domain of TFIIB is required for this interaction. The interaction between TAP and *tat* occurs at a 17-amino-acid (aa 34 to 48) contact site within *tat* activation domain (aa1 to 48). The contact sites for *tat* and TFIIB map within the TAP C-terminal region which contains the TAP activation domain (Yu et al., 1995a). It has been proposed that TAP is a cellular co-activator that bridges the *tat* transactivator to the general transcription machinery via TFIIB. More experiments are required to fully elucidate the mechanism of *tat* transactivation and to define the exact components of the preinitiation complex and that of the elongation complex in HIV-1 transcription.

1.2.6 *Rev* mechanism of action

Rev controls the early to late stage transition in viral RNA splicing. *Rev* is specifically involved in the protection and transport of full-length HIV-1 mRNAs to the cytoplasm. *In vitro* and *in vivo*, the key factor for the emergence of HIV-1 from early to late viral RNA synthesis following activation of the provirus is the presence of *rev* (reviewed by (Cullen and Greene, 1989)). In the absence of *rev*, long HIV mRNAs are retained in the nucleus by *cis* repressor sequences (CRS) located in *gag* and *env*. *Rev* is a 19-kDa nuclear protein that contains four functional domains : the basic domain (arginine rich domain: aa 34 to 50) binds to its RNA target sequence called Rev Responsive Element (RRE), which forms highly structured RNA stem-loops (Malim et al., 1989b). Two regions flanking this basic domain are required for multimerisation of *rev* on the RRE. Finally, five leucine residues near the C-terminus form the activation domain of *rev* (Malim et al., 1989a). It is proposed that *rev*/RRE interaction initiates with the binding of a *rev* monomer to a single high affinity RNA binding site (stem-loop IIB) within the RRE. This initial RNA-protein-interaction then induces the binding of additional *rev* molecules to adjacent, lower affinity sites leading to the formation of a multimeric complex. Southgate and colleagues (1990) first demonstrated that *tat/rev* fusion proteins can activate gene expression directed by an HIV-1 LTR promoter in which the TAR element has been replaced by an RNA sequence encoding the high-affinity binding site for *rev* (Southgate et al., 1990). The above system segregates *rev* RNA binding from *rev* function and hence provides an elegant assay for sequences of *rev* that are required for RNA binding and can also provide a means to examine the efficiency of *rev*

multimerisation (Madore et al., 1994). Madore and colleagues showed that *tat/rev* fusion proteins made with *rev* mutants (mutations in the *rev* basic domain, i.e., the RRE binding site) were not able to transactivate an HIV-1 LTR promoter in which the TAR had been replaced by the RRE stem-loop IIB (SLIIB). Mutations in *rev* outside the basic domain, but in other essential domains of *rev* did not affect the trans-activation activity of the fusion proteins on the RRE-LTR. These results showed that only mutations in the RNA binding motif of *rev* block its specific recognition of the SLIIB RNA target. This assay has been used to investigate the effects of specific mutations of *rev* on protein multimerisation *in vivo.* The same investigators showed that providing wild type *rev* in *trans*, re-established the trans-activation capacity of *tat/rev*^{mut} fusion protein. This rescue was achieved by the recruitement of *tat/rev*^{mut} protein to its promoter target after protein/protein interaction between the wild type *rev* (*rev*^{wt}) provided in *trans* and the mutant *rev* of the fusion protein. The complex *tat/rev*^{mu}/*rev*^{wt} can bind to the SLIIB of the RRE-LTR through the intact RNA binding domain of the wild type *rev.* By substituting *rev* mutants in place of the wild type, the authors have defined the *rev* domain involved in *rev* multimerisation as sequences flanking the basic motif of *rev* coupled with the *rev* activation domain. However, the basic domain of *rev* is dispensible for multimerisation. The different effects of identical *rev* mutations on *rev* multimerisation *in vitro* and *in vivo* (Malim and Cullen, 1991; Zapp et al., 1991), have supported the hypothesis that a cellular cofactor facilitates multimerisation *in vivo* while *in vitro* assays only measure the intrinsic ability of *rev* to multimerise on the RRE.

The precise mechanism of action of *rev* is not fully elucidated: *rev* might itself transport viral RNAs which contain the RRE, alternatively, since the basic domain

of *rev* also inhibits RNA splicing *in vitro*, *rev* could affect RNA splicing with its effects on RNA transport being secondary. Both mechanisms are not mutually exclusive. A better understanding of the *rev* mechanism of action was reached with the isolation of the murine cellular protein YL2 which modulates effects of *rev* in co-transfection assays (Luo et al., 1994). YL2 has 92% homology to the human p32 protein, which binds tightly to alternative splicing factor SF2/ASF that can commit pre-mRNAs to the splicing pathway (Fu, 1993). The human p32 interacts with the basic domain of rev and therefore, important for its effects on RNA splicing *in vitro*. Luo and colleagues (1994) proposed p32 as a co-factor of *rev* present in large mixed complexes (RRE/multimerised rev/p32) which not only inhibit splicing by sequestration of SF2/ASF but also stabilise unspliced mRNAs.

From the data generated so far, it appears that the formation of an active *rev*/RRE ribonucleoprotein complex *in vivo* may require an array of specific protein/protein and protein/RNA interactions that involve not only *rev* and the RRE but one or more cellular factor(s).

1.2.7 Kinetic phases of gene expression

The temporal aspect of HIV replication has been investigated (Kim et al., 1989) using single-cycle growth conditions for HIV in human lymphocytes. Under these conditions, the sequence of events is as follows:

1- Within 2-3h of viral infection, viral DNA is present in the cytoplasm.

2- Integrated DNA can be detected by 24h.

3- Initial products of HIV gene expression are the short, multiply spliced RNA species of 1.8 to 2 kilobases in length. These RNAs encode *tat, rev* and *nef.*

4- Other early mRNAs include the 4.3 kilobases singly spliced mRNAs coding for *env* as well as *vif* and *vpr.*

5- By 36 h, full-length transcripts (9.3 kb) which act as both genomic RNAs and mRNAs for the *gag-pol* polyprotein and for the *env* protein, have accumulated.

1.2.8 Production of *gag* and *pol* proteins by ribosomal frameshifting

HIV-1, like many retroviruses, makes the use of a translational frameshifting to overcome the termination codon at the end of *gag*. In HIV-1, one ribosomal frameshift event (-1) allows the synthesis of the fusion protein *gag-pol* in a ratio of 20:1. Reil and co-workers (1993) have shown that a slippery sequence of HIV-1 present in *gag*, consisting of a heptanucleotide with the sequence UUUUUUA, is sufficient *in vivo* to mediate a basal level of frameshifting (Reil et al., 1993). A stem-loop structure downstream of the slippery sequence in HIV-1, serves as a positive modulator in the process. The HIV-1 frameshifting mechanism is in accordance with the simultaneous slippage model, which proposes that the frameshift takes place at the slippery sequence and is stimulated by the secondary structure (Kollmus et al., 1994). The distance between the two *cis*-acting elements has been shown to play an important role in the efficiency of ribosomal frameshifting for a number of retroviruses, in the case of efficient slippery sequences UUUUUUA, as in HIV-1, the distance between the slippery sequence to the stem-loop does not need to be stringently maintained; it may vary from 3 to 9 bases (Kollmus et al., 1994).

1.2.9 Viral morphogenesis

Viral morphogenesis occurs at the late stages of HIV replication and involves the transport of viral proteins, *gag-pol* and *env* to the plasma membrane where *gag* and *gag-pol* precursor polyproteins assemble into virions and bud from the cell surface

in which membrane-anchored *env* proteins are present. *Gag* and *gag-pol* proteins form the internal core structure. HIV Gag is initially translated as a precursor Pr55, which is co-translationally modified by removal of the N-terminal methionine and attachment of myristic acid to the second N-terminal aa, glycine. During or more probably, after virus budding, the Pr55 Gag precursor is proteolytically processed by the *gag-pol* encoded protease (PR) into (from N- to C-terminal ends) matrix (MA; p17), (CA;p24), nucleocapsid (NC;p7) and p6.

The MA protein retains the myristate moiety and forms a matrix under the viral membrane; CA is the major core protein; NC contains the Cys-His motif required for viral RNA packaging and the p6 domain may be involved in the process of budding. The capsid is assembled from *gag* and *gag-pol* precursor polyproteins, in an approximate ratio of 20:1. As mentioned above, each polyprotein contains a myristic acid moiety at its N-terminus, which may serve to target the polyprotein to the membrane and/or keep the polyproteins anchored to the plasma membrane. Capsid formation is a poorly understood process, but presumably involves *gag/gag* and *gag/gag-pol* dimerisations and higher order multimerisation events. During capsid formation, two copies of the retroviral genomic RNA are selectively packaged into the capsid interior. The regions of the genomic RNA and *gag* polyprotein that interact to mediate packaging of the RNA are gradually being elucidated using genetic and biochemical analyses (reviewed in (Linial and Miller, 1990)). More recent work (Berkowitz and Goff, 1994); using RNA gel mobility shift assays with various segments of HIV-1 genomic RNA in combination with mutant *gag* polyproteins and the NC protein has shown that the strongest binding occurred with a 120 nt RNA molecule consisting of sequences flanking the gag-start codon, including three potential stem-loop structures. Two non-overlapping regions of NC,

each containing a Cys-His box, mediated specific binding to HIV-1 RNA. It is important to note that binding of HIV-1 RNA to NC proteins involves both secondary and tertiary structures formed by the RNA; hence the optimal binding might occur with the most stable stem-loop structures. The stabilisation of these structures and the folding in the correct conformation to allow binding to the NC proteins may also require specific flanking sequences in addition to the binding elements themselves. It is not known if the three stem-loop structures alone (the binding elements of HIV-1 RNA) are sufficient to promote encapsidation of the viral genomic RNA by the *gag* polyprotein *in vivo*. A more recent study analysing *gag* particle assembly and p17 dimerisation of *gag* mutants suggests that dimerisation of the *gag* precursor occurs by the interdigitation of alpha helices on adjacent matrix molecules (Morikawa et al., 1995).

Another important step in the HIV-1 particle maturation is the formation of biologically active envelope glycoprotein which involves post-translational cleavage of the precursor gp160 into gp120 and gp41 (McCune et al., 1988). Furin, a subtilin-like mammalian endoprotease, is thought to be responsible for the processing of gp160 of HIV-1, which share the consensus processing site motif (cleavage site), Arg-X-Lys/Arg-Arg, for protease recognition (for review see Barr, 1991; Nagai, 1993). The above notion has been questioned by Ohnishi and colleagues who showed that a Furin-defective cell-line (LoVo) is able to correctly process HIV-1 gp 160 (Ohnishi et al., 1994). LoVo cells were able to process HIV-1 gp160 of three different HIV-1 env genes with the same primary cleavage site sequence: Arg-Glu-Lys-Arg. When LoVo cells expressing gp160 were co-cultured with CD4 expressing HeLa cells, syncytia were formed showing that LoVo cells process gp160 to produce the biologically active fusion glycoprotein. The investigators also showed that LoVo

cells are able to produce infective progeny virions since co-transfection of LoVo cells with an HIV-1 DNA clone, resulted in the production of infectious virions. The above findings indicate that if furin is involved in gp160 processing it can be replaced with a different protease or that the processing of gp160 is mediated by protease(s) other than furin. One candidate is another mammalian subtilin, PACE 4, which displays, like furin, widespread tissue distribution and is presumed to participate in pre-protein processing. Another candidate is a 26kD protease which has been isolated from a human T-cell line and correctly processes gp160 *in vitro* (Kido et al., 1993).

1.3 HIV-1 VARIABILITY

Genetic variation of retroviruses is the combined result of the mutation rate per replication cycle, the number of replication cycles, and the fixation rate of mutations (i.e., the selective advantage or disavantage possessed by the variant virus; (Coffin, 1990)). A characteristic of reverse transcriptase is its lack of proof-reading ability which makes it highly error-prone compared to other DNA polymerases. The high error rate of HIV-1 RT is the major factor in the generation of mutations in the HIV-1 genome. The infidelity of HIV-1 RT has been studied with purified enzyme. Error rates have been found to be 5 to 10 per round of replication of the HIV genome (Preston et al., 1988). Moreover, it has recently been shown that HIV-1 RT copies both DNA and RNA template ends (before the jumps) with extremely low fidelity, yielding large amounts of product (30 to 50% of molecules processively polymerised) with 1 to 4 extra nucleotides added beyond the template termini. This non-templated polymerisation is highly specific for the addition of purine nucleotides (A > G) and could be an important mechanism for retroviral mutation (Patel and Preston, 1994). In addition, RT mediates strand-transfer reactions that lead to recombination. This results in the production of mutant proviral DNA and potentially

any of the protein components that make up new virions may differ to some extent from the parental strain. Thus, RT-catalysed polymerisation errors and recombination are major determinants of HIV-1 variability (Hu and Temin, 1990). It has been reported that these substitutions are not random in their distribution throughout the genome, with 'hot spots' where errors are more likely to occur (Roberts et al., 1988). All the above studies on purified enzyme, estimate the error rate of reverse transcriptase to be in the range of 5×10^{-4} to 6.7×10^{-4} mutations per bp per cycle, when determined for a DNA template and 2.5 x lower for an RNA template. Mansky and Temin (1995) have investigated the mutation rate of HIV-1 reverse transcriptase *in vivo* with the use of HIV-1 shuttle vectors (viruses carrying deletions in *gag-pol* and *env* but containing the *neo*gene, origin of replication, and the *lac*Z alpha peptide gene) (Mansky and Temin, 1995). These vectors replicate in mammalian cells as viruses and can be selected with the neomycin analog G418. The vectors can also replicate as plasmids in *E. coli* and are selected with the drug kanamycin. To be packaged into a virus particle, these vectors are complemented in *trans* by transient transfection of cells with an HIV-1 *gag-pol* expression plasmid and an amphotropic murine leukemia virus *env* expression plasmid. This system was used to study the mutations that occur during a single round of HIV-1 replication. The *in vivo* mutation rate was lower than that of purified enzyme, being 3.4×10^{-5} mutations per bp per cycle. The results also confirmed the *in vivo* observation that the most commonly detected mutations were base substitution mutations G to A and C to T transitions, and frameshift mutations (-1 frameshifts in runs of T's and A's) (Wain-Hobson, 1993). The other step where virions may acquire mutations is during transcription of the proviral DNA by RNA polymerase II.

Finally, the other potent mechanism for the generation of variability is probably genetic recombination. Recombination is a common property of retroviruses, for example in the AKR mouse model for leukaemogenesis; the development of leukaemia has been associated with the formation of new viruses by recombination between at least two different endogenous viruses (Chattopadhyay et al., 1982). Recombination can occur between the genomic RNAs, since retroviruses have a diploid genome, and particles can be heterozygous. Another form of recombination can occur at the time of integration, when retroviruses acquire sequences from the cell itself (eg. when retroviruses acquire oncogenes). This latter form of recombination has not been reported for HIV. Mixed infection with two distinct mutant viruses (each with a single base substitution in the RT gene conferring AZT resistance) passaged *in vitro* have resulted in the detection of recombinant progeny virions (Kellam and Larder, 1995). As much as 38% of screened clones were recombinant viruses after only 3 passages (Kellam and Larder, 1995). In the case of HIV-1, genetic recombination can be very efficient, especially in the presence of syncytium-inducing (S-I) variants where cell fusion occurs with the co-infection of different strains.

HIV, like other primate immunodeficiency viruses, displays greater variability than any other RNA viruses, and it is now clear that this feature is not a consequence of a more error prone reverse transcriptase (Mansky and Temin, 1995), but rather due to their ability to regulate their own gene expression, their powerful transactivation systems, and at least in the case of HIV-1, its high turnover rate *in vivo* involving continuous *de novo* replication (Ho et al., 1995; Wei et al., 1995). The main selective forces that have been proposed to drive HIV diversity are the immune response, cell tropism and random activation of infected cells. The end result is that

in HIV-1 infection, multiple genetic variants exist at the same time within one infected host. This population of virus variants is referred to as a quasispecies (this heterogenicity of the virus is also termed 'population polymorphism').

The genetic variability of HIV-1 influences its ability as a pathogenic agent, for example, variants that differ in replication rate and cell killing capacity (virulence), may alter the severity of the disease and the rate of progression. In this respect, the importance of the appearance of syncytia inducing (SI) strains for disease progression has been clearly demonstrated and will be discussed in more details in section 1.3.1. Genetic variation can lead to variants able to escape from the specific humoral and cell-mediated immune controls (see Section 1.3.2), variants with altered tropism for various cell types, or both (McKnight et al., 1995). Finally, the appearance of variants with decreased sensitivity to different antiviral drugs also occurs (Larder, 1994).

1.3.1 Syncytia- and non-syncytia-inducing variants

The demonstration of phenotypic changes of HIV-1 during the course of disease has preceded that of genotypic changes (Asjo et al., 1986). In early HIV infection, viruses isolated have been referred to as slow/low, replicating slowly *in vitro*, with low p24 levels. In contrast, viruses isolated from infected individuals with advanced disease have been referred to as rapid/high, replicating more rapidly *in vitro* with a cytopathic phenotype (Asjo et al., 1986; Cheng-Mayer et al., 1988; Fenyo et al., 1989). Subsequently, Tersmette and co-workers showed that this heterogeneity of viral isolates can also distinguish HIV strains that grow well in macrophages or T cell lines. Many macrophage-tropic viruses neither directly kill cells nor do they induce syncytia. In contrast, the T cell replicating viruses are highly cytopathic and syncytia-

inducing. These two groups of viruses have been termed NSI and SI (non syncytia-inducing and syncytia-inducing) viruses (Tersmette et al., 1988; Tersmette et al., 1989a). Phenotypic variants are viruses exhibiting altered growth characteristics as a result of changes in the viral genome. Genotypic studies have concentrated on the region of the genome coding for the envelope proteins, particularly on V3, the third of the five hypervariable regions in the gp 120 portion of the envelope gene. The V3 domain has been shown to contain determinants of virus cytopathicity, cell tropism, and virus infectivity.

Sequential analysis of V3 genomes from HIV-1 infected individuals as well as studies on the biological properties of V3 chimeric viruses have led to a defined picture of the mutations determining these properties. The mutations in V3 causing a transition from NSI to SI occur 2 years before disease development and are substitutions of basic amino acid residues at positions 306, 320, and 324: S to R, D to Q to R, and D to N respectively (Fouchier et al., 1992; de Jong et al., 1992a; Kuiken et al., 1992). Syncytium formation occurs when infected cells expressing envelope protein bind other cells expressing CD4 receptors, forming multinucleated giant cells or syncytia. The predominance of NSI or SI variants appears to depend upon the disease status of the infected individual. However, several investigators have shown that at seroconversion and in early infection, a high degree of sequence homogeneity in *env* is found, and that strains are of the NSI phenotype (Pang et al., 1992; Wolfs et al., 1992; Cichutek et al., 1992b; McNearney et al., 1992; Ait-Khaled and Emery, 1993; Zhang et al., 1993; Zhu et al., 1993). A mechanism for the counter-selection against SI viruses has been proposed by Zhu and colleagues based upon a study of the genotypes and phenotypes of the viruses present in five seroconverters and two of their corresponding sexual partners (Zhu et al., 1993).

The seroconverters harboured a homogeneous viral quasispecies which was macrophage tropic and NSI. In contrast, the two transmitters harboured viruses with high degree of sequence heterogeneity and a spectrum of different phenotypes. The gp 120 sequences of the transmitted viruses were found to match best with minor, rather than major, variants in the blood of the transmitters. The proposed explanation for these findings is the selective transmission of one viral variant that is able to penetrate more effectively the mucosal barrier of the new host. The observation that the transmited viruses were macrophage-tropic and NSI, suggests that their selective penetration is a result of efficient replication in macrophages or antigen-presenting cells present in the submucosal space. However, from what is now known on the distribution and sequestration of HIV-1 variants in different organs (Donaldson et al., 1994a; Delassus et al., 1992a; Fox et al., 1991) and see Chapter 6, it would have been interesting to analyze the HIV-1 variants in the semen of the transmitters. Moreover, this model cannot be applied in the case of hemophiliacs or intravenous drug users infected through blood to blood transmission (where there is no mucosal barrier to cross) and where this counter-selection against SI variants has also been demonstrated (Zhang et al., 1993). A different model has been recently proposed suggesting that the immune system of the virus donor may play the more important role in transmission. This hypothesis is based on the observation that virus variants found early in infection are macrophage-tropic, grow poorly in T cell lines, and, more importantly, are relatively resistant to antibody mediated neutralisation. Combining the above observations with the abundance of antibodies during most phases of HIV infection, it has been suggested that the majority of virions in an asymptomatic individual are coated with antibodies rendering them much less infectious. The macrophage-tropic viruses, being more resistant to neutralisation in the transmitter would have an increased chance of

establishing infection upon transmission. This model assumes that transmission is established by cell-free virus and that the majority of asymptomatic HIV-1 infected individuals have an effective autologous neutralising antibody response. Which role each of the proposed mechanisms plays in this transmission "bottleneck" that counter-selects against SI viruses remain to be elucidated. The selection in favour of NSI, macrophage-tropic viruses, would "reset the clock" and cause the evolution of HIV-1 quasispecies to begin anew in each newly infected individual.

Studies on the transition of NSI to SI phenotype during the course of disease have shown an association between the appearance of SI virus and a subsequent more rapid fall in CD4 cell count (Tersmette et al., 1989a; Tersmette et al., 1989b; Koot et al., 1993). Koot and coworkers followed a group of 188 untreated HIV-seropositive men over a period of 30 months. Over this period, 70.8% of the men with SI variants progressed to AIDS, compared to 15.8% of individuals without SI variants at entry. In this study group, the rate of CD4 cell decline was 3.7-fold higher in subjects who had SI variants at entry than in subjects with NSI variants throughout the follow-up period, with an intermediate rate of decline for the 22 individuals who converted to SI viruses during the study. Acceleration of the course of disease which follows the appearance of SI variants is probably brought about by the two different characteristics of the viral strains.Firstly, the marked difference in cell tropism: NSI virus being monocytotropic and SI virus being T-cell tropic. Secondly the SI viruses are characterised by a much faster viral replication than NSI viruses *in vitro*. This increased replication rate would increase the likelihood of mutations occurring and, hence, the probability that mutant viruses with selective advantages emerge (escape mutants, anti-viral resistant mutants, or cell-type adapted mutants). Not all patients who progress to ARC or AIDS harbour SI viruses;

some patients carry NSI variants with an accelerated rate of replication. However, a strong correlation has been shown between the rate of CD4 cell count decline and virus replication rate in sequential isolates from an individual (Figure 1.9). When the transition to SI virus occurs, a rapid depletion in CD4 count is observed and opportunistic infections begin within 6 to 18 months. Despite this clear association , the cause-effect relationship is difficult to establish. The presence of SI virus was assessed in 325 individuals and correlated with CD4 cell decline and clinical end-points (Richman and Bozzette, 1994). Findings similar to those of the Dutch group were reported: adjusted for differences in demographics, treatments, baseline CD4 count, beta2-microglobulin and p24 antigenemia, the mean rates of CD4 cell decline were 40 and 102 cells/year in the NSI and SI groups respectively. The rates of CD4 decline in 16 persons converting from NSI to SI virus averaged 31 cells/year before conversion and 142 cells/year after conversion ($p=0.04$). The study also demonstrated reduced survival rates in patients with the SI phenotype, which was mainly attributable to the impact of the SI phenotype on the CD4 cell decline. In other words, when controlling for the CD4 cell count, the SI phenotype did not predict the immediate risk of death. In summary, the presence of SI strains predicts an accelerated decline in CD4 cell count but CD4 cell absolute number, regardless of how quickly the reduced count is attained, is of better prognosis value for disease progression (Richman and Bozzette, 1994). The more rapid decline of CD4 cell counts in individuals infected with the SI phenotype of HIV-1 may be the *in vivo* reflection of the *in vitro* characteristics of SI variants, which are to induce multinucleated giant cells (syncytia) and cell death in PBMC and in T cell lines such as MT-2, whereas, NSI isolates can grow to titers as high as 10^6 infectious units/ml of culture without producing microscopic cytopathology.

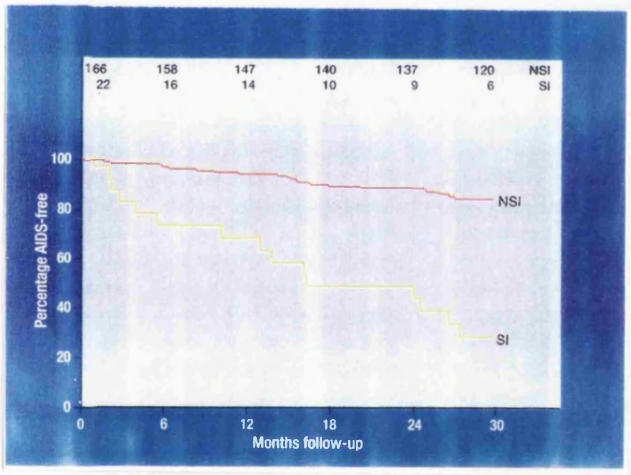


Figure 1.9: Kaplan-Meier plot of the cumulative progression to AIDS of 188 untreated men. Subjects were grouped according to their NSI or SI status at entry. Reproduced from Boucher and Larder, 1994.

A more recent study by Simmonds and collaborators (Donaldson et al., 1994a) argues against the above logical proposal for explaining the association between SI variants and accelerated CD4 cell decline. Their study provides data suggesting that the *in vitro* SI phenotype should not be over-interpreted and extrapolated to the *in vivo* situation. These investigators have analysed the *in vivo* distribution of variants and cytopathology of HIV-1 in post-mortem tissue samples from 11 autopsied HIV-infected patients. The V3 loop sequences of all patients were of the NSI/ macrophage tropic phenotype as determined by the charge and amino-acid composition of the sequences, irrespective of the stage of disease. Immunocytochemical staining with anti-p24 antibodies revealed the major cell types productively infected in each organ. The predominant expressing cells in the lung and the brain of the AIDS patients were macrophages and microglia (in brain), frequently forming multi-nucleated giant cells (syncytia) even though the V3 loop sequences of these variants resembled those of the NSI isolates *in vitro*. Contrary to the other studies, the results of Simmonds and colleagues argue that variants with the SI/T-cell tropic phenotype *in vitro* are not essential for disease progression or *in vivo* cytopathic effects. In this context, it is important to note that isolates with the *in vitro* phenotypes of NSI/ macrophage tropic or SI/T-cell tropic are both able to grow in PBMCs and primary macrophages. The high frequency of variants with a predicted NSI phenotype in the above study is not inconsistent with previous studies. Variants with predicted NSI phenotypes (based on genotype) are frequently detected in other published analysis of viral sequences *in vivo*, in many cases from patients with advanced HIV disease (Epstein et al., 1991; Holmes et al., 1992; Keys et al., 1993; Kuiken et al., 1992; Milich et al., 1993; Wolfs et al., 1992; Wolinsky et al., 1992) and see chapter 4, although their frequency relative to SI variants varies considerably between patients.

1.3.2 Antigenic variants

The vigorous HIV-specific immune response can be a selective force *in vivo,* driving HIV-1 variability. It has been suggested that there may be a continuous cycle in which HIV mutates in T and B cell epitopes to escape immune recognition followed by renewed immune responses to these mutated epitopes (McCune, 1991). There is some experimental evidence for this process in the humoral immune response to HIV: neutralising antibodies in HIV-infected individuals can neutralise the initial HIV isolates but not subsequent isolates (Reitz, Jr. et al., 1988; Watkins et al., 1993), and analysis of serial isolates of HIV from two infected individuals has revealed the evolution of strains with a mutation in the V3 loop (aa 306) at a site critical for antibody binding (Wolfs et al., 1991). It is important to note that resistance to neutralization has been mapped within or outside of the epitope identified by monoclonal antibodies which shows that conformational changes of the envelope influence antibody recognition. Moreover, mutations in V2 and gp41 have also been shown to confer resistance to neutralization (for review see Levy, 1994). These findings on sensitivity to antibody neutralization emphasize the influence of other envelope domains on the overall conformation of the viral envelope and hence effective antibody binding. A more recent study, investigated the existence of neutralisation escape variants (NEV) in patients after seroconversion to determine if there was a correlation between disease progression and the emergence of NEV with altered cytopathogenicity and cell tropism (Tsang et al., 1994). In the above study, the investigators followed 5 subjects between 2 to 5 years after primary HIV-1 infection. In 3 subjects, the initial virus isolate from seroconversion could be neutralised by autologous serum, but isolates obtained at two subsequent time points exhibited reduced sensitivity to serum neutralisation, decreased replication in primary macrophages, and increased ability to induce syncytia. Two of these 3

individuals progressed to AIDS and died. On the other hand, sequential isolates from the other two subjects showed variability in sensitivity to serum neutralization and biological features. The clinical status of these patients remained stable. This study confirmed that viruses isolated at seroconversion appear to be NSI, macrophage-tropic, and that reduced macrophage tropism and SI phenotype are correlated with disease progression, but is the first to associate the emergence of NEV from the time of primary infection with disease progression. There is also evidence for selection of escape mutants from CD8+ cytotoxic T-lymphocytes(CTL) (Phillips et al., 1991; Rowland-Jones et al., 1992).

Antigenic variation presents a considerable problem for vaccine development as it is essential to induce broadly cross-protective immunity against the diverging strains of HIV-1 across the world. Antigenic variation has also been proposed to account for HIV pathogenesis. Initially, the host raises an excellent immune response to HIV which leads to the long asymptomatic period following seroconversion. Antigenic variation permits escape from the immune system, and the resurgence of sequential virus variants could lead to AIDS. This hypothesis has been investigated by Nowak and May (Nowak and May, 1991) in a mathematical model that can fit all disease outcomes (acute disease, delayed disease, and no disease). The authors argue that the generation of antigenic variants causes an asymmetric interaction between immunological and viral diversity. While each virus strains impairs immune responses by diminishing MHC-II-restricted T-helper cells, strain-specific immune responses can only control specific virus variants. Eventually, a diversity threshold is reached when the immune system fails to control the virus population. This model suggests that HIV infection is an evolutionary process that determines the time scale from infection to disease. However, this hypothesis is not supported by the observed

disease course in the animal models available. Indeed, the genetic diversity and load of SIV appears to be similar to HIV, yet only humans develop AIDS. Moreover, chimpanzees can be infected with HIV-1 and immune escape mutants occur, but they do not develop disease. Therefore, it is unlikely that a threshold of antigenic diversity alone causes disease. The detailed 55 month follow-up of two patients (p007 and p020) CTL responses to defined epitopes by McMichael and collaborators revealed two different patterns. In patient 007, who remained well, only a single reactive epitope was detected. Upon sequencing, two variants of the epitope were detected, and CTLs reactive against each variant were also present. In contrast, patient 020, who died of AIDS, harboured CTLs reactive against three epitopes in a temporally fluctuating pattern. Interestingly, sequence variants of each of the epitopes also fluctuated over time (Phillips et al., 1991). The pattern observed suggested a possible oscillation of the immune response to specific epitopes driven by virus escape mutants at the various epitopes. A recent mathematical model of the CTL response has been constructed to investigate the factors responsible for different patterns of cellular immune response (Nowak et al., 1995b). The model predicts a stable immune response when the initial response is mounted against an immunodominant epitope which remains homogeneous, in which case a equilibrium between the cells displaying the epitope and its specific CTL will be reached. This stable response can only be sustained if the virus remains homogeneous at the dominant epitope; if escape mutants of adequate replicative ability do arise, the oscillatory pattern occurs with responses shifting from one immunodominant epitope to another. These immune responses require help from CD4+ T cells, therefore as CD4+ T cells decline the cellular arm of the immune system weakens to eventually being entirely unable to generate new CTLs against appearing virus variants (Figure 1.10).

1.3.3 Neurotropic variants

Neuropathogenic strains of HIV-1 have been characterised to be biologically, serologically and molecularly distinct from peripheral blood isolates (Anand et al., 1987; Cheng-Mayer et al., 1989; Watkins et al., 1990). In general, the CNS-derived isolates, although growing to high titre in primary CD4+ lymphocytes, are not cytopathic for these cells. In addition, brain isolates grow to higher titres in macrophages than blood isolates. The CNS isolates can rarely be propagated in established T- B- or monocyte cell lines. These observations suggest that distinct but related viral strains could evolve independently in the infected host. Most genetic studies of CNS isolates and natural variants have focused on the *env* region of HIV-1 genome and have demonstrated mutations in V3 determine the tropism of CNS variants for monocyte derived cells (Takeuchi et al., 1991). The differences in cell tropism determined by the envelope sequence could be explained by the affinity of the envelope molecule to the still unidentified accessory receptor(s) required for HIV-1 entry in its target cell.

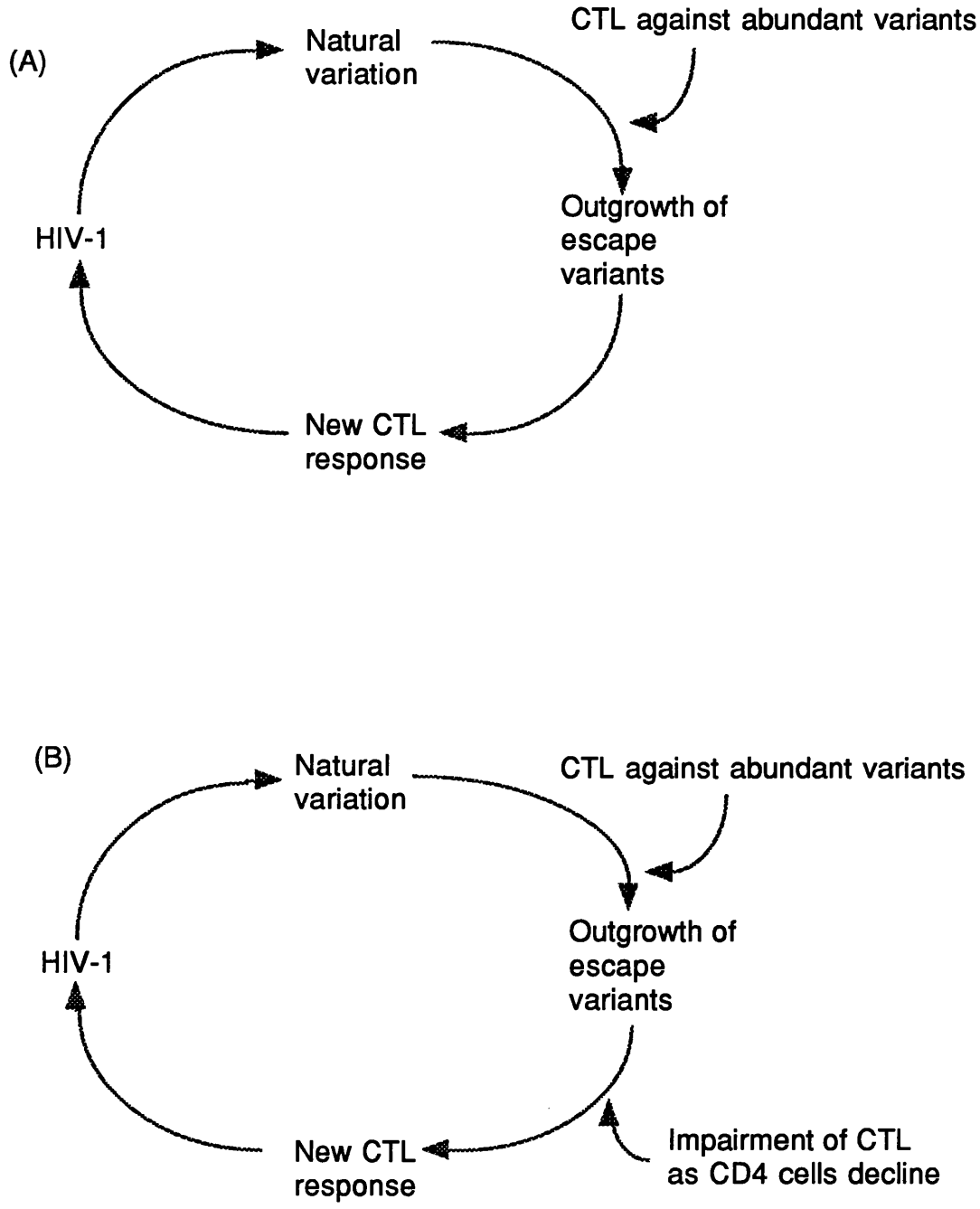


Figure 1.10: Model of CTL escape variants and disease progression based on the data from Nowak et al., 1995b. (A) Early to intermediate stage of disease. (B) Late stage of disease.

Despite the fact that most studies on tissue tropism have focused on the envelope gene of HIV-1, they do not exclude the possibility that other regions of the virus genome may be implicated in tissue tropism and adaptation. A region of particular interest is the LTR which controls HIV-1 replication under the influence of specific cellular transcription factors and whose potency would depend on the prevalence of particular combinations of cellular factors in different cell types (see section 1.3.4).

1.3.4 HIV-1 LTR variation *in vivo*

The majority of studies on the effect of HIV LTR sequence variation on its transciptional efficiency have used laboratory-adapted or laboratory-engineered molecular clones (see sections 1.2.4 and 1.2.5). By contrast, studies on natural LTR genotypes are scarce. A detailed analysis of a single patient over four years by Delasssus and co-workers (Delassus et al., 1991; Delassus et al., 1992b) reported numerous point mutations, which did not affect basal-level transcription. However, the different genotypes displayed a wide range of transactivation levels when tested by transient transfection of reporter vectors into SW480 cell (Delassus et al., 1992b). Moreover, longitudinal analysis of the TAR/*tat* genotypes from this patient over four years (during which time the patient progressed to a symptomatic stage) showed no temporal enrichment of more transcriptionally active genotypes (Delassus et al., 1992b). Another study by Golub and co-workers, has also reported the isolation of two HIV-1 isolates, derived from a single parental isolate, with different replicative capacities. The virus with the better replicative capacity in chronically infected cells had a more potent LTR than the slower replicating isolate as demonstrated by LTR-directed CAT gene expression. This increased basal activity of the LTR was due to the combined effect of two mutations: a point

mutation at position -94 affecting the four-nucleotide spacer between the two NF-κB sites and a 24bp duplication between positions -128 and -151 (Golub et al., 1990). This duplication overlaps the LTR TCF-1alpha binding site (see section 1.2.5.5). Englund and co-workers reported the isolation of three HIV-1 molecular clones derived from a blood quasispecies with distinct replicative properties in MT-4 cells. The phenotypic differences observed between the 3 clones were attributable to changes in the LTR potencies which upon sequencing revealed polymorphisms in the NF-κB sites (Englund et al., 1991). Length polymorphisms have also been reported in the LTR region at the TCF-1α and Sp1 sites in 5 of 17 noncultivated patient samples by Koken and colleagues (Koken et al., 1992). These investigators showed that the variant with 4 Sp1 sites outgrew wild-type virus (3 Sp1 sites) when equal amounts of the size-variant and wild-type molecular clones were co-transfected into T-cell lines. The most extensive study on LTR variation has been reported by Michael and co-workers (Michael et al., 1994), and consisted of a longitudinal analysis of molecular clones from the LTR/*gag* leader region amplified directly from the peripheral blood of four patients over three years. The investigators reported a large number of point mutations and length polymorphisms in the regulatory elements of the HIV LTRs. Length polymorphisms were associated with duplications of Sp1 and TCF-1α sites but not of the NF-κB sites. These mutations conferred a wide range of transcriptional activities to the LTR when assessed by using reporter gene analysis. This study showed that mutations and duplications have complementary effects on LTR potency. For example, it was shown that the increased transcriptional activity of variants containing four Sp1 sites only ocurred in LTRs without TCF-1α duplications. No generalised trend in transcriptional activity of the LTRs was observed in the four patients (Michael et al., 1994). However, it is important to bear in mind that functional assays investigating transcription potencies

of LTR variants in reporter genes systems involves the use of different cell types, different mechanisms to deliver the *tat* protein in the cells, and does not take into account the genomic background of the LTR analysed. All of these factors limit the interpretation of the above studies. Thus, the comparative analysis of *in vivo* potencies of LTR variants is very difficult to establish, and requires the isolation of the full HIV variants in a number of primary human cells suceptible to HIV infection, together with a assessment of the replication kinetics of these variants.

1.4 PHYLOGENETIC ANALYSIS OF HIV-1

1.4.1 Geographic diversity of HIV-1 strains

The global molecular epidemiogy of HIV-1 has been extensively studied, and there is now a wealth of HIV sequence data from all over the world forming a database which is constantly updated by reseach groups worldwide. Phylogenetic analysis of HIV-1 sequences, primarily on the basis of *env*, and *gag* regions of isolates from diverse geographic locations in the world has identified at least eight distinct subtypes or clades, designated subtypes A to H (Myers, 1994). Each of these subtypes is approximately equidistant from the others in a phylogenetic pattern described as a star phylogeny (Korber et al., 1994). Most recently, the genetic analysis of HIV-1 strains from 24 patients in Cyprus has identified a new subtype designated subtype I (Kostrikis et al., 1995). In this study, the gp120 sequences from two of the 24 patients formed a distinct clade upon phylogenetic analysis with representative sequences of all other subtypes. This subtype designation based on the branch lengths and topology of the tree was approved by the Los Alamos National Laboratory (Figure 1.11). Many studies similar to the one above, as well as identifying new subtypes, have also evaluated the prevalence of these subtypes in

different geographical areas. These studies allowed important epidemiological findings to be drawn, for example, determining if single or multiple introductions of a particular subtype occurred in an area with a recent epidemic, and the implications of such data for the global evolution of HIV-1 and future vaccine development.

Recently Charneau and co-workers proposed dividing the HIV-1 isolates into two groups, group M (major), and group O (outlier). The latter group comprises isolates from Cameroon (ANT70, MVP5180, and VAU), that have only 64 to 70% similarity to other HIV-1 subtypes, and isolates from this subgroup also have an atypical western-blot reactivity (Louwagie et al., 1995; Gurtler et al., 1994; Loussert-Ajaka et al., 1994; Janssens et al., 1994; Nkengasong et al., 1994; Zekeng et al., 1994; Charneau et al., 1994). Group O isolates are also very distant from chimpanzees isolates such as one isolated from a chimpanzee captured in the wild in Gabon, namely CPZGAB (Figure 1.12).

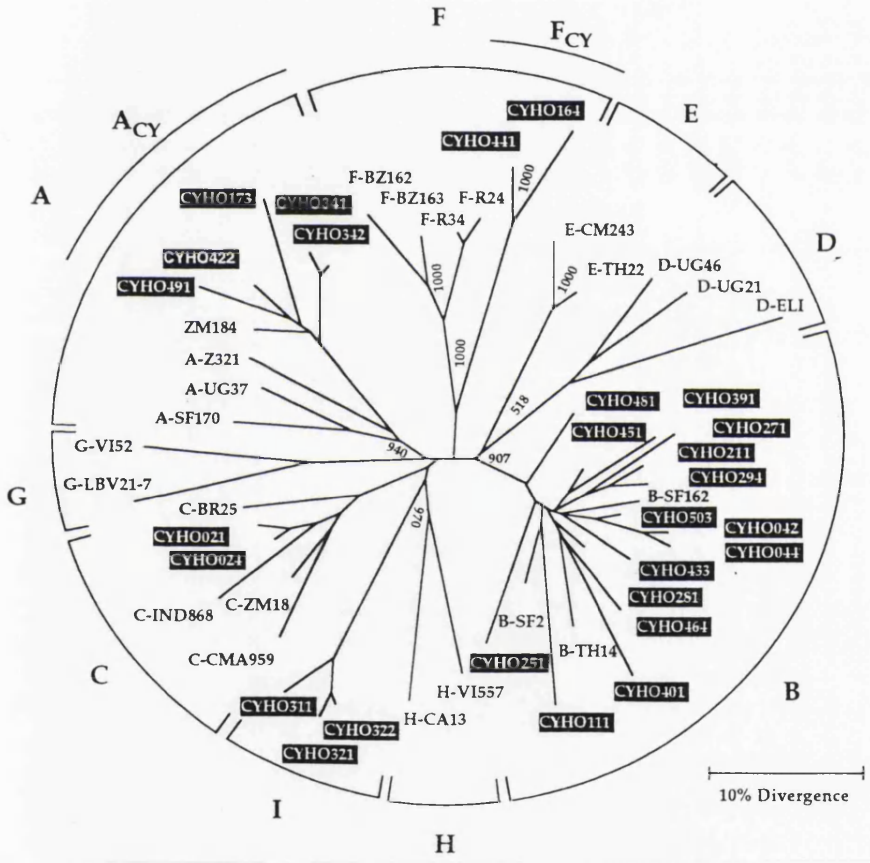


Figure 1.11: Neighbor-joining phylogenetic tree relating 27 HIV-1 gp120 sequences from patients in Cyprus (Kostrikis et al., 1995) and 23 representative reference sequences from eight characterised subtypes (A through F). The sequences determined by Kostrikis and colleagues are shown in white on black boxes.

The fact that isolates from group O are less common than those from group M in the pandemic, and appear to be strictly geographically linked to Cameroon and Gabon would suggest that group O is:

1) a recently evolved variant or alternatively

2) is a variant with a low pathogenicity and/or infectivity.

A study by Loussert-Ajaka and colleagues (1995) has analysed the variability of HIV-1 group O strains isolated from 7 Cameroonian patients living in France (Loussert-Ajaka et al., 1995) and showed that the genetic distances based on *gag* and *env* C2V3 regions between group O isolates were comparable to those observed in M intersubtypes, demonstrating that the group O viruses are genetically very diverse and may thus have a comparable age to the M group viruses (Figure 1.12). This study, therefore, does not support explanation *1)* (see above) to account for why Group M dispersed first, and is the dominant lineage in the pandemic. The pathogenicity and natural history of these viruses is under investigation and will take several years to establish. However, it is clear that group O viruses are associated with AIDS, since, amongst the individuals reported to be infected with group O strains only, two infected patients have died of AIDS, and four others have reached CDC stage IV (as of October 1995). It has been suggested that epidemiologic factors (i.e., population of relatively low mobility) may account for the delayed emergence of group O HIV-1 strains.

The extraordinary potential of HIV-1 to evolve at a "fastforward" rate must be taken into account in HIV diagnosis and in vaccine and drug development.

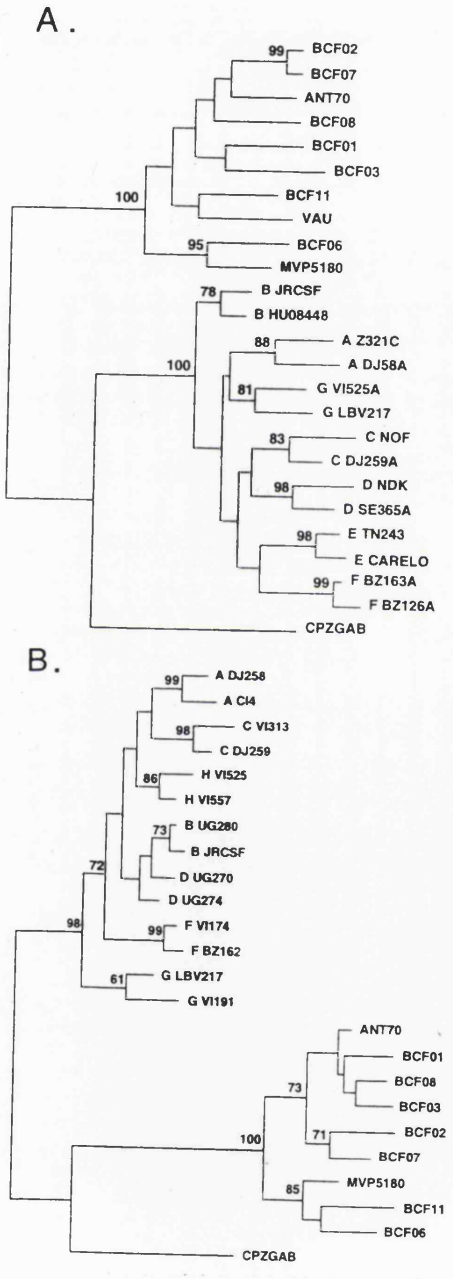


Figure 1.12: Weighted parsimony phylogenetic trees relating (A) the HIV-1 *env* C2 to V3 region and (B) CAp24 *gag* coding sequences (Loussert-Ajaka et al., 1995).

1.4.2 HIV quasispecies

The global variability of HIV-1 described above illustrates the rapid evolution of the virus. The rate of this evolution is in part determined by the generation of variants in the infected hosts. Indeed, the environment of continuous *de novo* virus infection, extensive replication and high CD4 turnover provides an ideal setting for the generation of HIV variants. There are a vast number of studies demonstrating that a single individual harbours many related HIV-1 variants at a single time point referred to as a quasispesies. These studies have investigated intra-patient evolution by analysing HIV-1 variants present in sequential samples during disease progression (Cheynier et al., 1990; Delassus et al., 1991; Delassus et al., 1992b; McNearney et al., 1992; Fouchier et al., 1992; Cichutek et al., 1992a; Pedroza Martins et al., 1992; McNearney T et al., 1995); variants harboured at seroconversion and in donor-receiver pairs (Zhang et al., 1993; Zhu et al., 1993; Pang et al., 1992; Ait-Khaled and Emery, 1993; Wolfs et al., 1992; Cichutek et al., 1992b; Cichutek et al., 1991), variants present in mother-child samples (Wolfs et al., 1992; Wolinsky et al., 1992; Wike et al., 1992; Scarlatti et al., 1993; Ahmad et al., 1995; Mulder-Kampinga et al., 1995; Briant et al., 1995), variants present in different organs of infected individuals (Ball et al., 1994; Donaldson et al., 1994a; Ait-Khaled and Emery, 1994; Ait-Khaled M et al., 1995) and finally variants present in distinct regions (individual white pulps) of a single spleen (Delassus et al., 1992a; Cheynier et al., 1994). Using phylogenetic analysis, which is a display of the relatedness of variants, the investigators of the above studies have shown that HIV-1 variants can evolve independently in different organs, it is hypothesised that the differences between variants reflect cell tropism and cell-adaptation in multiple organs. In contrast, Simon Wain-Hobson who pioneered these detailed studies of changes in time and changes in space of HIV-1 quasispecies in the infected host cautions

against the over interpretation of these findings derived from a restricted number of clones of relatively small regions of the genome and supports the hypothesis that HIV-1 diversity is relatively neutral, simply generated by random activation of latently infected T-cell clones. Indeed, calculations based on the assumption that an HIV infected individual will have between $5x10^6$ and $5x10^9$ infected lymphocytes depending on disease stage and that the point mutation rate per genome per cycle for HIV is between 0.1 and 1, he concludes that an HIV-1 infected individual will harbor between $5x10^5$ and $5x10^{10}$ genetically distinct variants.

It is clear that the generation of progeny virions is random, but there are several selection forces driving HIV evolution such as the constraints on structure and function of proteins and nucleic acids (negative selection), the founder effect, ie, the bottleneck transmission of a small number of variants, the requirement for rapid growth of virus in a particular cell type (including reverse transcription, transport of the proviral DNA to the nucleus, integration and viral transcription), and the host immune system. Each region of the HIV-1 genome is exposed to the above selection forces to different extents.

1.5 NATURAL HISTORY OF HIV-1 INFECTION

Although the course of HIV infection varies among individuals, a common pattern of development has been recognised. Primary infection with HIV is followed by the development of detectable cellular then humoral immune responses to the virus and a prolonged period (median 10 years) of clinical latency. During this variable period the patient is usually asymptomatic. Subsequently, clincally apparent disease develops,ie, the manifestations of profond immunodeficiency, and even with treatment, death usually occurs within 2 years.

Following infection with HIV-1, approximately 50% of people experience an acute illness characterised by fevers, chills, rash, lymphadenopathy and sometimes neurological symptoms (Cooper et al., 1985). At this time, virus levels in the plasma are high, and virus can easily be cultured from peripheral blood and CSF (Daar et al., 1991). This high viral burden in the periphery is controlled by a strong humoral and cellular immune response; HIV specific cytotoxic T cells have been detected as soon as few days after infection, and their appearance is concomitant with a decreased viral load (Pantaleo et al., 1994; Koup et al., 1994).

1.6 PATHOGENIC MECHANISMS OF HIV DISEASE

The natural history and pathogenic process of HIV-1 infection in humans are complex and variable, depending on a multitude of viral and host factors and their interactions. However, it is now clear that HIV-1 infection is characterised by continuous virus replication counteracted by a strong immune response during the variable period of "clinical latency", a battle eventually lost by an exhausted immune system (Ho et al., 1995; Wei et al., 1995). Viruses can cause disease by several mechanisms (pathogenesis). These include direct virus induced cell lysis, immunopathology, apoptosis, and indirect virus toxicity (soluble viral proteins), and which of these is the main determinant of pathogenesis depends upon the virus under consideration, the organ in which the pathology is observed, and possibly the stage of disease (temporal association). The slow but constant decline of CD4 cells despite the low viral burden in peripheral blood mononuclear cells during the clinical latency of HIV-1 infection had initially led the scientific community to favour indirect mechanisms of pathogenesis whereas the more recent data highlighting the high viral burden throughout HIV infection supports a more direct role of HIV in the CD4 decline. The remaining question is whether HIV infection leads to CD4 depletion by

HIV-induced cytolysis (either directly and/or through fusion-induced syncytia formation) or by an immunopathological process, such as an antiviral cytotoxic T-lymphocytes (CTL) immune response. The latter mechanism is argued by some investigators on the basis that HIV per se is non-cytopathic (Zinkernagel and Hengartner, 1994). However, the better understanding of the correlates of protective immunity to HIV infection/disease highlights the beneficial role of a strong cell-mediated immune response arguing against an immunopathological mechanism of disease (Haynes et al., 1996). There are experimental data in support of both theses (direct viral cytolysis or immunopathology) (Safrit and Koup 1996; Dalgleish 1996), but the relative importance of either as the cause of CD4+ T- cell decline susequent to HIV infection remains to be determined. It is also important to note that the relative contribution of one or the other mechanism may significantly differ depending on the stage of HIV infection (primary infection, "clinical latency", or late stage of disease) under consideration.

As mentioned above, recent studies investigating viral burden in lymphoid organs using sensitive molecular techniques have extended and refined the initial reports based on electron microscopy and immunocytochemical methods, and have shown that the lymphoid organs are the major reservoir of virus and preferential site of replication. Pantaleo and collaborators (Pantaleo et al., 1991) using PCR and reverse transcriptase PCR, have shown that 5 to 10 times more HIV-1 infected cells and a higher concentration of both regulatory and structural mRNAs are present in the lymphoid organs than in peripheral blood. This viral burden in lymphoid organs is greater than that in blood throughout the period of clinical latency. In early and intermediate disease, HIV particles complexed with antibodies and complement accumulate in the lymphoid organs where they are trapped in the network of

follicular dendritic cells present in the germinal centres. This subclinical lymphadenopathy may also be caused by the altered circulatory kinetics of CD4 T cells, including those infected with HIV, as these may be unable to enter the circulation. As the late stages of the HIV disease evolve, the architecture of the lymph nodes is disrupted characterised by the involution of the FDC network and lymphocyte depletion, resulting in the loss of the trapping capacity and the free recirculation of virions and HIV infected cells. The above findings have been interpreted to account for the massive increase in peripheral viral burden in late-stage disease (Pantaleo et al., 1993a). This high frequency of CD4 T lymphocytes and macrophages infected by HIV during the incubation period of AIDS has also been demonstrated by Embreston and collaborators who used *in situ* PCR and *in situ* hybridisation to analyse lymph node biopsies (all of which showed follicular hyperplasia and normal germinal centres) from 4 HIV-1 infected individuals (Embretson et al., 1993).

Similar findings have been reported for SIV infection of rhesus macaques. Of particular interest is the elegant work of Chabraski and coworkers who investigated the dynamics of spread of SIV early after infection and thereafter, using RNA *in situ* hybridisation (Chakrabarti et al., 1994a,b,c). The results showed that the infected macaques had a phase of acute viral replication in lymph nodes between 7 to 14 days post-inoculation which paralleled that observed in blood, followed by a second phase characterised by a controlled productive infection limited to the lymph nodes where viral particles were trapped in the forming germinal centres. This study, conducted on 8 macaques, showed that even though the peaks of productive infection were similar in all animals, the number of productively infected cells that persited in the lymph node after primary infection was variable between animals. In

contrast, no difference was observed in the blood compartment where productive infection was efficiently controlled. The 1- to 2-year follow-up of the macaques has revealed that the degree of viral persistence in the lymph nodes was directly associated with disease progression (Chakrabarti et al., 1994a). These findings indicated that the course of primary SIV infection in lymph nodes is variable, which suggests that the initial capacity of the host to control productive infection in lymph nodes is an important determinant of the rate of disease progression. The above results are in accordance with findings reported by Pantaleo and colleagues (1995) who investigated the histopathological, virologic and immunologic characteristics of subjects with long-term nonprogresssive HIV infection defined as individuals with 7 or more years of documented HIV infection, with more than 600 CD4+ T cells/μl , no anti-retroviral therapy, and no HIV associated illness (Pantaleo et al., 1995). The study involved 15 subjects with long-term nonprogressive HIV infection and 18 with progressive HIV disease. The results showed that lymph nodes from nonprogressors had significantly fewer hyperplastic features, no lymphocyte depletion, and none of the involuted features characteristic of subjects with progressive disease. Furthermore, plasma levels of HIV-1 RNA and the proviral burden in the peripheral blood mononuclear cells were both significantly lower (4 to 20 times) in nonprogressors than in individuals with progressive disease. These results are in accordance with those reported in a recent study demonstrating that the concentration of plasma viral RNA after primary HIV-1 infection is predictive of the long term clinical outcome (Mellors et al., 1995). Moreover, Pantaleo and colleagues showed that HIV could not be isolated from the plasma of nonprogressors but virus was consistently cultured from the mononuclear cells from the lymph nodes demonstrating viral replication. The degree of virus trapping in the follicular dendritic cells (FDC) network was lower in the nonprogressors and

paralleled the extent of germinal centre formation. The authors proposed that this lower degree of trapping reflects the relatively nonreactive state of the lymphoid tissue and may contribute to the lower viral load and hence lower rate of tissue activation (absent or small germinal centres) observed in subjects with long-term nonprogressive infection. The study also showed the presence of high titres of neutralising antibodies (as tested with HIV-1_{IIIB} and HIV-1_{MN} , not with primary isolates) and consistent detection of HIV-specific cytotoxicity. Overall, subjects with long-term nonprogressive HIV infection had preserved lymphoid tissue with reduced formation of germinal centres and reduced HIV trapping, concomitent with a low but persistent level of viral replication.

A more recent finding demonstrated a detrimental role of FDC virus trapping (Heath et al., 1995) in allowing antigen presentation to immune effector cells leading to the initiation and propagation of an immune response to the pathogen. However, in the case of HIV, this trapping may have a detrimental effect on the host by infecting CD4 T cells which reside in the paracortical region of lymphoid tissue and migrate into the B cell-rich germinal centres to provide help for B cells . Indeed, Heath and co-workers have shown that HIV-1 trapped in the FDC web was highly infectious, and that FDCs can convert neutralised HIV into an infectious form even in the presence of vast excess of neutralising antibodies. This study suggests that FDCs may be the mechanism by which HIV infection can occur efficiently even in the presence of neutralising antibodies, and provides further evidence for the hypothesis proposed by Pantaleo and co-workers (see above).

Another major advance in the understanding of HIV infection has come from two important studies reported by the groups of Ho and Wei (Ho et al., 1995; Wei et al.,

1995). Both groups have investigated the dynamics of HIV production by analysing the changes in plasma viral RNA loads and CD4 cell levels in HIV-1 infected patients during treatment with potent inhibitors (viral protease inhibitors and a non-nucleoside RT inhibitor, Nevirapine). Treatment of patients resulted in a rapid reduction of free virus in the plasma within a few days, with a half life of approximately 1.8 days in both the above studies. Knowing that blood viremia is constant over several weeks in untreated patients, the investigators assumed that the viral clearance equals viral production to provide an estimate that between 10^8-10^9 HIV virions are produced and destroyed per day. The decrease in viraemia during treatment was paralleled with a CD4 cell recovery. The CD4 cell recovery rate in the periphery was used to estimate that 2×10^9 cells are destroyed and produced everyday. However, this latter conclusion has been disputed by several groups since the CD4 cell recovery observed in the periphery may merely be a redistribution of CD4 T cells from the lymphoid tissues (the mobilization hypothesis) rather than new production from thymic progenitors or from the proliferation of peripheral CD4 T cells (the proliferation theory), see scientific correspondence (Nowak et al., 1995a; Mosier, 1995; Lai et al., 1995; Weber and Galpin, 1995; Michie, 1995; Duesberg and Bialy, 1995; Buianouckas, 1995; Ascher et al., 1995; Bukrinsky et al., 1995; Phillips et al., 1995; Dimitrov and Martin, 1995; Sprent and Tough, 1995). In defence of their thesis, Ho's group have argued that the surface-marker phenotypes of CD4 lymphocytes post-therapy were activation markers indicative of lymphocyte repopulation by post-thymic cellular proliferation rather lymphocyte redistribution.

All the above studies on viral burden and dynamics of replication clearly show that HIV "...is replicating 24 hours a day and from day one...." (Wain-Hobson, 1995) and therefore provides a strong rationale for therapeutic strategies targeting the virus as

early as possible after infection (before or during seroconversion). Indeed, the first controlled trial of anti-retroviral treatment (AZT monotherapy at diagnosis of seroconversion and for 6 months thereafter) in primary HIV-1 infection showed that such intervention improved the susequent clinical course and increase the CD4 cell count at the end of the 36 months follow-up (Kinloch-De Loes et al., 1995).

1.7 AIMS OF THE PROJECT

The overall aim of this study has been to define the variability of HIV-1 LTR and V3 region loci in infected individuals at multiple body sites as well as the changes of this variability during disease progression.

Aim 1

To investigate the genetic relatedness of HIV-1 LTR natural proviral variants present simultaneously in the PBMC and lymph node of infected individuals. In the light of growing evidence that the lymphoid tissues are the major reservoir of virus in HIV infection and the privileged sites of continuous replication at all stages of disease, it is pertinent to hypothesise, that being the site of replication, lymphoid organs may also be the predominant site of variation and evolution of HIV-1 quasispecies.

Aim 2

To investigate the phylogenetic relationtionship between HIV-1 proviral variants present in sequential blood samplesand those present in the lymph node. The temporal fluctuation of HIV-1variants is also analysed.

Aim3

To assess the nature of HIV-1 V3 region variants present in the PBMC of a patient who was undergoing HIV-1 seroconversion and and compare it to that of the variants present in the lymph node and PBMC of a patient with advanced disease.

Aim 4

To investigate the possible role of HIV-1 LTR as a determinant of cell-type adaptation by analysing the phylogenetic relationship of HIV-1 proviral LTR variants present in postmortem samples of lymphoid tissues, lung, and nervous tissues of an HIV-1 infected patient who died with AIDS.

CHAPTER 2

MATERIALS AND METHODS

2.1 INTRODUCTION

This chapter summarizes all the methodologies used in the present study. The study focused on the longitudinal and cross-sectional analysis of HIV-1 sequence variation. The analysis has been carried out on clinical samples obtained from five HIV-1 infected individuals. The summary of the clinical course of the 5 patients is presented, it includes: patients 0, and 1 (chapter 4), patients 1, 2 and 3 (chapter 5) and patient 4 (chapter 6). This chapter also describes the clinical specimens obtained from the 5 patients, the DNA extraction procedures, the amplification of HIV-1 proviral DNA, the cloning of the amplicons into pUC18, the sequencing method and the computing methods used for sequence variation analysis.

The technique that has allowed a major advance in the field of molecular biology in the late 1980s has been the polymerase chain reaction (PCR). PCR is an *in vitro* method for primer directed enzymatic amplification of a specific DNA target. Two oligonucleotides which flank the DNA target sequence are used to prime the activity of a DNA polymerase enzyme. This pair of oligonucleotides or "primers" specific to a sequence of interest are submitted to repeated cycles of heat denaturation, primer annealing and extension of the annealed primers by a DNA polymerase enzyme. During each of these cycles the amount of DNA present in the previous cycle is potentially doubled leading to an exponential increase (2^n, n being the number of cycles) of the DNA target sequence originally present in the reaction. The first description of DNA amplification used the *Klenow* fragment of DNA polymerase I for primer extension, but denaturation of the double stranded target DNA which requires temperatures above 90°C also denatures the enzyme, necessitating the addition of new *Klenow* at every cycle (Saiki et al., 1988). The use of thermostable DNA polymerases, such as that of the bacterium *Thermus aquaticus* (*Taq*), which

remains stable and enzymatically active at high temperatures, have been crucial in the development of the PCR technology. The main benefit of this development was to limit the manipulation of the reaction mixture during the procedure and therefore reduce the possibilities of contamination and shorten the process time. Indeed, PCR is an extremely sensitive technique and contamination with target DNA not originally present in the sample analysed has been the main problem in the use of PCR as a research and diagnostic tool (see chapter 3).

The commercially available *Taq* polymerases are highly purified or recombinant forms of the natural enzyme, which lack endonuclease and 3' to 5' exonuclease activities but retains their 5' to 3' polymerisation-dependent replacement activity. *Taq* polymerase is active at 72°C, a temperature which has the benefit of allowing high specificity of the PCR by dissociating non-specific annealing of primers, therefore allowing only primers bound to their specific complementary sequence to remain intact for the extension reaction.

The conditions of the PCR are crucial for the success of a specific and sensitive amplification. In this chapter , the PCR conditions for the proviral DNA amplification of two regions of the HIV-1 genome, namely, the LTR and the V3 region are described (see chapter 3 for optimisation of LTR PCR). V3 data are presented for patients 0 and 1 only (see chapter 4), LTR data are presented for patients 1, 2, 3, and 4 (see chapters 5 & 6).

2.2 PATIENTS ANALYSED

2.2.1 Patients clinical course and clinical specimens obtained

Patient 0: C.O (see chapter 4)

The individual analysed presented at day 0 with diarrhoea and rash (no blood sample was available for analysis). Thirteen days later, after exclusion of bacterial pathogens, HIV seroconversion illness was suspected and antibodies to HIV-1 (Abbott EIA) found to be equivocal. At the same time HIV antigens (Abbott) were positive (15pg/ml). Two days later, HIV antigens became negative and antibodies to HIV-1 were strongly positive. The patient had elevated $ß_2m$ levels (3.3mg/l), a depressed CD4% (18%), and a CD4/CD8 ratio of 0.9 (normal range 0.7-3.3). A PBMC sample in preservative free heparin was obtained from the patient at day 13 after presentation.

Patient 1: I.R (see chapter 4 and 5)

A PBMC sample and neck lymph node sample were obtained in December 1992. The patient had been on zidovudine for 3 years and had a CD4 count of 220 x 10^6/l. Pneumocystis carinii pneumonia and lymphoma were diagnosed. Histology showed follicular lysis and disruption of the germinal centres of the lymph node. A blood sample was obtained 14 months prior to the lymph node biopsy, the CD4 count at that time was 270x 10^6/l. Finally, a blood sample was obtained 6 months following the biopsy and the CD4 count was 110x 10^6/l.

Patient 2: D.H (see chapter 5)

A PBMC sample and mediastinal lymph node biopsy were taken in December 1990. The patient had suspected lymphoma which was not confirmed but infection with *Mycobacterium Avium Intracellulare* was detected. The absolute CD4 count at biopsy was 18 x 10^6/l. A subsequent PBMC sample was taken from the same individual in November 1991. The absolute CD4 count at this time was 7 x 10^6/l. Histology of the lymph node biopsy showed germinal centres to be intact except for a small epithelial cell aggregate.

Patient 3: C.L(see chapter 5)

A PBMC sample and axilliary lymph node sample were obtained in May 1992. The patient had received no antiretroviral therapy and had a CD4 count of 680 x 10^6/l. Toxoplasma myocarditis and extrapulmonary TB (lymph node) were diagnosed. Histology showed extensive destruction and replacement of the lymph node by granulomas. A second PBMC sample was obtained 17 months after the lymph node biopsy, by which time the CD4 count had declined to 420 x 10^6/l

Patient 4: A.A (see chapter 6)

The patient first presented in January 1990 with a 4 week history of cough and dyspnoea, weight loss, sweating and rash and was subsequently diagnosed HIV-1 antibody positive. *Pneumocystis carinii* and cytomegalovirus were isolated from bronchial lavage washings. The patient said that the only occasion he had been at risk of acquiring HIV was six weeks before his first admission. The absolute CD4 count at this time was 378 x 10^6/l. At discharge the patient was prescribed zidovudine (200 mg 5 times a day) and nebulised pentamidine (300mg monthly).

The patient was subsequently treated for a succession of pathological conditions: cytomegalovirus (CMV) pneumonitis (1/90) requiring ventilation, cutaneous Kaposi Sarcoma (KS) (9/91), CMV oesophagitis (1/92), pleural effusion (2/92), pulmonary KS and CMV in broncho-alveolar lavage (BAL) (10/92). The patient started an alternating regimen (every 2 months) of zidovudine and didanosine (200mg twice daily) in April 1992 and died in November of the same year. A peripheral blood sample (10ml) in preservative free heparin was obtained in September 1992, 2 months prior to death. Full autopsy was undertaken by Dr James McLaughlin from the Department of Histopathology at the Royal Free Hospital and the following tissues were further investigated: spleen, lung, lymph node, spinal cord and dorsal root ganglion.

2.2.2 Determination of markers of HIV-1 infection in peripheral blood

Antibodies to HIV-1

Antibody tests were performed by the virology department, Royal Free Hospital, UK. Antibodies to HIV were measured by enzyme immunoassay according to the manufacturer's instruction ('Wellcozyme', Wellcome Diagnostics, Dartford, UK).

Beta-2-microglobulin

A commercial radial immunodiffusion (RDI) method was used to measure β-2-microglobulin in serum samples (NanoRDI, Binding Site Ltd, Birmingham, UK) and were monitored by QC controls and the UK NEQAS quality assurance scheme.

HIV-p24 core antigen, acid dissociation assay (Du Pont)

Monitoring HIV-p24 antigenaemia is valuable for providing diagnosis of congenital HIV transmission and to identify seroconverting patients. An acid dissociation is included in the assay protocol in order to disrupt antibody/antigen complexes in patients' sera which are not detectable by standard ELISA assays. The dissociation is achieved by using a combination of low pH and heat after which the samples are neutralised and run in the HIV-1 p24 core ELISA assay. The ELISAs were performed by the Virology Department, Royal Free Hospital, London.

2.2.3 Determination of T-cell subsets

Absolute CD4 and CD8 lymphocyte counts were performed in the Department of Immunology, Royal Free Hospital, London. Absolute lymphocyte counts were determined by an automated whole blood counter (Ortho 'ETL 800') with differential screen. A whole blood method was used, and the percentage of CD4 and CD8 lymphocytes analysed by flow cytometry using a FACScan (Becton Dickinson, Crawley, UK). A monoclonal antibody, RFT4, to the CD4 antigen, and a CD8 monoclonal antibody, RFT8, to the CD8 antigen were used throughout, in double concentration with a monoclonal CD3 antibody (OKT3 or UCHT1). Flow cytometer quality control was carried out using Q.C. beads, and by the UK National External Quality Assurance Scheme (NEQAS).

2.2.4 Summary of CDC classification of HIV infection

HIV-1 infection is characterised by a long clinical latency followed by the appearance of symptoms when the immunodeficiency of the infected individual is profond. The symptoms and opportunistic secondary infection subsequent to immunodeficiency occur in stages and are used to classify HIV-infected individuals in disease categories of increased severity. This classification (by the center for disease control (CDC), USA) is summarised in table 2.1.

Table 2.1: Summary of CDC classification of HIV infection

I Acute infection (seroconversion)
II Asymptomatic infection
III Persistent generalised lymphadenopathy

IV A Constitutionai disease
One or more of : fever>1 month, involuntary weight loss> 10% of baseline, diarrhoea>1 month (in absence of causes other than HIV)

IV B Neurological disease
One or more of : dementia, myelopathy, peripheral neuropathy (in absence of causes other than HIV)

IV C Secondary infectious disease (symptomatic or invasive)

IV CI Opportunistic infectlons

Pneumocystis carinii pneumonia	Cryptococcosis
Chronic Cryposporidiosis	Histoplasmosis
Toxoplasmosis	Mycobacterial infection
Extra-intestinal Strongyloidiasis	Cytomegalovirus infection
Isosporiasis	Chronic Mucocutaneous Herpes Simplex infection
Herpes Simplex Ulcer>1 month	Extrapulmonary Tuberculosis
Oesophageal, Bronchial or pulmonary candidiasis	Progressive multifocal Leukoencephalopathy

IV C2 Others

Reccurent Salmonella Bactereamia	Nocardiasis
Pulmonary Tuberculosis	Multidermatomal Herpes Zoster
Oral Hairy Leukoplasia	Oral Candidiasis

IV D Secondary cancers
Kaposi Sarcoma; Non-Hodgkins Lymphoma (Small non-cleaved Lymphoma or immunoblastic sarcoma); Primary Lymphoma of the brain

IV E Other conditions
eg: neoplasms other than those in **D**, or reccurent infections other than those in **C**.

2.3 DNA EXTRACTION FROM CLINICAL SAMPLES

2.3.1 Extraction from Peripheral Blood Mononuclear Cells (PBMCs)

2.3.1.1 Preparation of PBMCs for use or storage

All mixtures were prepared in sterile conditions and stored at 4°C.

Mixture A: RPMI medium containing:

Penicillin (100 U/ml)

Streptomycin (100mg/L)

Mixture B: RPMI medium containing:

Penicillin (100 U/ml)

Streptomycin (100mg/L)

Foetal calf serum (30% v/v)

DMSO (10%)

All the equipment used for the procedure were also sterile, the processing of blood was carried out in a category 2 safety cabinet.

1- 10 ml of preservative free heparin blood were centrifuged at 400g for 10 min

2- The plasma was removed in 2ml aliquots and stored at -70°C

3- The cells were resuspended in PBS to a final volume of 20 ml, and were well mixed by inversion.

4- The cell suspension from step 3 was carefully layered onto 15 ml of Ficoll-paque in a 50ml Falcon tube.

5- Centrifuge at 400g for 30mins at 20°C.

6- The upper layer was pipetted off and discarded.

7- The layer formed by the PBMCs at the interface was removed with pastette, keeping the volume of cells and Ficoll below 10ml.

8- The cells were washed with 40ml of mixture A by gentle resuspension, and centrifuged at 400g for 10 min.

9- The cell pellet was resuspended in 50 ml of mixture A for a second washing, and then centrifuged at 400g for 10min.

10- The cells were finally resuspended in 2 ml of Mixture B (freezing medium), aliquoted in 2 X 1ml aliquots.

11- The aliquots were placed in a polystyrene box at -70°C overnight for slow freezing, and then transferred for permanent storage. The cells were then used for DNA extraction.

2.3.1.2 DNA extraction from PBMCs

DNA was extracted and purified from PBMCs by proteinase K digestion, phenol extraction and ethanol precipitation using established methodologies. In brief, 5 million cells were resuspended in 1 ml of 0.5% SDS, 10 mM Tris-HCl pH7.5, 1mM EDTA, and 500µg/ml of proteinase K, and incubated for 2 to 4 hours at 37°C. Total DNA was extracted once with 1/2 volume of phenol, once with an equal volume of phenol/chloroform, and once with an equal volume of chloroform. The DNA was precipitated by adding 1/10 volume of 4M sodium acetate and 3 volumes of 100% ethanol to the aqueous phase of the chloroform extraction. The DNA was precipitated at -70°C for at least 1 hour. After centrifugation at 13 000g for 20 min, the DNA pellet was washed in 70% ethanol (500µl), re-centrifuged at 13 000g for 5 min, and then vacuum dried for 5-10 min. Finally the DNA pellet was resuspended in 100µl of TE buffer adapted from Sambrook et al., 1989. The DNA was quantified

both by taking optical density measurements and/or agarose gel visualisation (see Section 2.3.3).

2.3.2 Extraction from post-mortem tissues

Each tissue ($1cm^3$) was finely dissected and washed 3 times in 1 ml PBS to remove residual blood. The samples were then washed 3 times for 2 minutes in 1 ml of 10 mM Tris-HCl (pH 7.5), 1mM EDTA and 0.5% SDS. The tissues were then incubated overnight at 37°C in 1ml of extraction buffer: 10mM Tris-HCl pH8, 10mM EDTA, 1% SDS and 50µg/ml proteinase K.

Cellular DNA was extracted twice with equal volume of phenol and twice with phenol/chloroform. The resulting aqueous phase was precipitated in 2.5 volumes of ethanol and 1/10 volume of 4M NaCl at -70°C for at least 2 hours. After centrifugation at 13 000g for 20 minutes, the pellet was washed in 200µl of 70% ethanol and then dried. Finally, the pellet was resuspended in 100µl of sterile distilled water. The DNA was quantified both by taking optical density measurements and/or agarose gel visualisation (see Section 2.3.3).

2.3.3 Quantitation of Nucleic Acids (DNA/RNA)

DNA concentrations were accurately measured by spectrophotometric absorbance readings of diluted DNA and RNA solutions at 260nm and 280nm. An absorbance of 1.0 at 260nm was taken to be equivalent to a concentration of 50µg/ml for double-stranded DNA, and 40µg/ml for single-stranded DNA and RNA. For oligonucleotides (amplimers) an OD of 1 corresponds to ~20µg/ml. The purity of the preparations was estimated by the ratio of the absorbances at 260nm and 280nm. Pure DNA and RNA preparations have an OD_{260}/OD_{280} ratio of $\geq$ 1.8 and 2.0, respectively. The RNA

contamination of the extracted DNA was estimated upon electrophoresis, which showed negligible amount of degraded RNA (low molecular weight), and should not affect the quantitation of the extracted DNA.

Ethidium bromide fluorescence upon UV illumination (302nm) is directly proportional to the amount of DNA in which it has intercalated. Therefore, a rapid and relatively accurate method of double stranded DNA measurement is the visual comparison of the intensity of DNA bands to the intensity of λ molecular weight marker bands containing a known amount of DNA.

2.4 AMPLIFICATION OF HIV-1 PROVIRAL DNA BY PCR

2.4.1 HIV-1 V3 PCR

PCR reactions were performed on extracted PBMCs' DNA (0.5 µg) using a Hybaid DNA thermal cycler in the following 10 x reaction buffer: 250 mM Tris (pH 8.4), 170 mM $(NH_4)_2SO_4$, 100 mM ß-Mercaptoethanol, 0.02% gelatin, and 40 mM $MgCl_2$ with 2 units of Taq polymerase. The cycling conditions are shown in table 2.2. The primers were obtained from the MRC (Aids Directed Program (ADP) reagent repository numbers: ADP 826 and ADP 828). Primers for the reactions consisted of a (+) sense oligonucleotide spanning positions 6971 to 6997 and a (-) sense oligonucleotide spanning positions 7313 to 7343 , and were 5'-phosphorylated prior to use. The sequences were as follows:

5'-CGCTAGGAATTCGGCCAGTAGTATCAACTCAA,

5'-GTACACAAGCTTCTGGGTCCCCTCCTGAGGA.

The PCR amplicon was 373 bp in length (based on the HIV-1 HXB2 prototype sequence).

The PCR was optimised by varying the concentrations of $MgCl_2$ in the 10X Taq buffer and also by adjusting the annealing temperature of the PCR reaction. 10X Taq buffers were made to $MgCl_2$ concentrations of 20, 30, 40, 50, and 60 mM and PCR was performed with 150 ng of each primer. Both a negative and positive control were included (to test for false positive reactions and system failures respectively), and the PCR program was used with an annealing temperature of 60°C (Table 2.2). The optimum concentration of $MgCl_2$ was found to be 4 mM. The temperature of the annealing reaction was varied to determine the highest temperature permissible for the reaction to occur so that non-specific binding of primer to the target would be excluded.

Example of a PCR mixture:		
	10.0 µl	10X Taq buffer
	10.0 µl	forward primer (15 ng/µl)
	10.0 µl	reverse primer (15 ng/µl)
	3.0 µl	dNTP mixture
	0.2 µl	Taq polymerase (5 U/µl)
	5.0 µl	PBMCs' DNA (100 ng/µl)
	61.8 µl	SDW
	100.0 µl	Total

100 µl of molecular biology grade light mineral oil (Sigma) was overlaid on each 100µl PCR mixture.

Buffers:

10X Taq buffer:		
	250 mM	Tris-HCl, pH 8.4 (Sigma)
	170 mM	ammonium sulphate (BDH)
	50-100 mM	$MgCl_2$ (see text) (BDH)
	100 mM	β-mercaptoethanol (BDH)
	0.02% (w/v)	gelatin (BDH)

dNTP mixture:	6.25 mM	dATP (deoxyadenosine triphosphate)
(Amersham)	6.25 mM	dCTP (deoxycytosine triphosphate)
	6.25 mM	dGTP (deoxyguanosine triphosphate)
	6.25 mM	dTTP (deoxythymidine triphosphate)

2.4.2 HIV-1 LTR PCR

PCR reactions were performed on extracted cell DNA (0.5 µg) using a Hybaid DNA thermal cycler (see table 2.2). Primers consisted of a (+) sense oligonucleotide spanning positions -395 to -375 and a (-) sense oligonucleotide spanning positions +125 to +147 relative to the start of transcription, and were 5'-phosphorylated prior to use (see Section 2.5.1). The primers were obtained from the MRC (ADP reagent repository numbers ADP 810 and ADP 811). The sequences were as follows:

5'-CACACAAGGCTACTTCCCTGA,

5'-GATCTCTAGTTACCAGAGTCAC.

Table 2.2 PCR Program Used for the Amplification of the HIV-1 LTR and V3 region

Stage	Temperature (°C)	Duration of stage (mins)	Number of cycles
1.1	94.0	4.00	1
1.2	60.0	1.15	
1.3	72.0	1.15	
2.1	94.0	1.15	32
2.2	60.0	1.15	
2.3	72.0	1.15	
3.1	94.0	1.15	2
3.2	60.0	1.15	
3.3	72.0	10.0	

2.4.3 Decontamination of PCR mixture by UV irradiation

The use of UV irradiation for PCR decontamination was assessed in an independent study, the PCRs carried out on clinical samples for sequence analysis were not submitted to irradiation. Indeed a multiroom setting was available in the department and showed to eliminate any contamination problems. For the independent study (see above), irradiation was carried out using two wavelengths of UV light simultaneously, 302 and 254nm, delivered by a UVP transilluminator (model TM15; Genetic Research Instrumentation, Cambridge) and a UVP hand-held lamp (model UVGL-50; Genetic Research Instrumentation) respectively. All the reagents for the PCR, including primers and Taq polymerase, but excluding target DNA, were dispensed into 0.5 ml polypropylene eppendorf tubes (Alpha Laboratories) resulting in a final volume of 90µl per tube. Duplicate tubes were exposed to UV irradiation for 5 minutes by placing them directly onto the screen of the transilluminator, with the hand-held lamp positioned above the tubes, "sandwiching" them between the two UV sources, thus maximising the dose delivered to the PCR mixture. After UV exposure template DNA was added and the tubes placed in the thermal reactor. The effect of UV irradiation on the PCR sensitivity was investigated by comparing the detection threshold of the PCR system when the PCR reagents are subjected to irradiation before or after addition of target. The efficacy of UV exposure to decontaminate the PCR was investigated by carrying out the same procedure as above but adding a range of plasmid DNA template prior to UV exposure. Target DNA ranged from 10^6 to 5 copies of recombinant pUC18 plasmid with the HIV-1 LTR fragment inserted at the *Sma* I site of the polycloning region.

2.4.4 PCR visualisation on agarose gel

The presence of PCR products from each experiment were confirmed by 1.8% agarose gel electrophoresis in 1X TBE at 100 V for approximately 30 minutes.

Media used in agarose gel electrophoresis

10X TBE

432 g Tris base (Sigma)

220 g boric acid (Sigma)

37.2 g EDTA (Sigma)

made up to 4 litres with distilled and deionised water.

Agarose gel (e.g. for 1.8% w/v agarose gel)

50 ml of 1X TBE was added to 0.9 g of agarose and the latter was melted in a 750 W microwave oven for approximately 3 minutes at full power in a 250 ml conical flask. 0.5 µl of a 10 mg/ml ethidium bromide solution was added in a fume hood when the solution had cooled to approximately 50°C.

2.5 PREPARATION OF THE PCR AMPLIFIED HIV-1 LTR AND V3 REGION FOR CLONING

2.5.1 Phosphorylation of primers

In order to blunt-end ligate the amplicon into *Sma I* digested and dephosphorylated pUC18, the primers were phosphorylated prior to use in the PCRs. 10ug oligonucleotide primer, 1/10 volume of 10x Kinase buffer (0.5mM Tris-HCl pH 7.6,

0.1M $MgCl_2$, 50mM DDT, 1mM Spermidine, 1mM EDTA), 5µl 10mM ATP, 10 units of T4 polynucleotide kinase, and distilled water to a final volume of 30µl was incubated at 37°C for 30 min. The reaction mixture was then frozen in a º20 C freezer to stop the reaction.

2.5.2 Dephosphorylation of vector DNA

Sma I digested pUC 18 was purchased from Boehringer Mannheim (Germany). To prevent recircularisation of vector and allow blunt-end ligation with the 5' phosphorylated amplicons, the vector was dephosphorylated. 10µg of pUC18 was treated with 0.1 unit of calf intestinal alkaline phosphatase, 1/10 volume 10x One-Phor-All buffer pH 7.5 (10mM Tris-acetate, 10mM magnesium acetate, 50mM potassium acetate (Pharmacia)), adjusted to a final volume of 20µl with distilled water and incubated at 37°C for 30 min. The enzyme was inactivated at 85°C for 15 min. The vector DNA was purified by Phenol extraction, ethanol precipitation and resuspended in 10µl of distilled water, as described in Section2.3.2.

2.5.3 Purification of phosphorylated PCR product

The equivalent of 4 PCRs (320 ul) was purified by the GeneClean procedure (Stratagene) for use in the ligation reaction. Briefly, the GeneClean procedure involves excision of the DNA fragment from a low melting point agarose gel followed by addition of 1/2 volume of TBE Modifier and 4.5 volumes of NaI stock solution (Stratagene). The gel was melted at 55°C for 5 minutes and 10 µl of GLASSMILK suspension was added, mixed and incubated on ice for 5 minutes. The GLASSMILK was pelleted at 13 000g for one minute and the supernatant removed. The pellet was washed three times with 50 volumes of ice cold NEW WASH and the DNA was eluted from the matrix by incubation at 55°C for 3 minutes in 25 µl of TE (pH 8.0).

2.5.4 Klenow repair of the purified amplicon

Any potential recessed termini present in the PCR products (10µl of Gene Cleaned amplicon) were filled in by the addition of 2µl of 10XMSK buffer, 4µl of 2mM dNTPs, and 2 units of Klenow fragment of DNA polymerase I, in a final volume of 20µl. The reaction mixture was incubated at room temperature for 30 min, and then stored at -20°C until use for ligation reactions.

10X Medium salt Klenow buffer

13.2 mM Tris-HCl pH 7.4

13.2 mM $MgCl_2$

100 mM NaCl

2 mM DTT

2.6 LIGATION OF THE PCR AMPLIFIED HIV-1 LTR AND V3 REGION TO pUC18

10X ligation buffer

100 mM Tris-HCl pH 7.4

100 mM $MgCl_2$ (BDH)

100 mM dithiothreitol (DTT) (Sigma)

10 mM spermidine (Sigma)

10 µg/ml bovine serum albumin (BSA) (Sigma)

Ligation of the HIV-1 LTR and V3 amplicons within *Sma* I cut dephosphorylated pUC18 vector was performed as follows. Approximately 10 to 20 ng of purified PCR product was ligated to 75 ng of *Sma* I cut dephosphorylated pUC18 vector using 1.25 U of T4 DNA ligase (Northumbria), 1 mM ATP (Amersham) in 1X ligation

buffer in a 10 µl reaction. The ligations were incubated at 16°C for at least 4 hours or overnight. Competent JM105 cells were transformed with each ligation product (using the method described in section 2.7.2) with a background control (75 ng of *Sma* I cut dephosphorylated pUC18 vector). These transformed cells were plated onto LB agar plates (containing 50 µg/ml ampicillin, 40µg/ml X-gal, and 40µg/ml IPTG) and incubated at 37°C overnight (see Section 2.7).

2.7 PREPARATION AND TRANSFORMATION OF COMPETENT ESCHERICHIA *COLI* (E. *COLI*) BACTERIAL CELLS WITH THE RECOMBINANT pUC 18 PLASMIDS

In this section, buffers and procedures for the amplification and purification of plasmid DNA are described.

Media and buffers for plasmid preparation

Luria-Bertani (LB) broth

0.5% (w/v) NaCl (Sigma)

0.5% (w/v) yeast extract (Sigma)

1.0% (w/v) tryptone (Oxoid)

The pH of the broth was adjusted to pH 7.5 with NaOH sterilised by autoclaving.

Bacteriological agar (Oxoid) (1.5%) was added for LB-agar plates.

X-Gal

20mg/ml in dimethylformamide.

IPTG

200mg/ml in SDW.

Phenol:Chloroform

500 g phenol (Sigma)

70 g M-cresol (Sigma)

96 ml chloroform (BDH)

4 ml isoamyl alcohol (Sigma)

The mixture was equilibrated with three 200 ml aliquots of 0.1 M NaCl, 0.1 M Tris-HCl, pH 8.0 and stored at 4°C until use.

RNAse A Preparation

10 mg/ml RNAse A (Boehringer Mannheim)

10 mM Tris-HCl (pH 7.5) (Sigma)

15 mM NaCl

The solution was heated at 100°C for 15 minutes to remove DNAse activity and cooled slowly to room temperature. Aliquots were stored at -20°C.

Solution I

50 mM D-glucose (BDH)

10 mM EDTA pH 8.0 (Sigma)

25 mM Tris-HCl pH 8.0

The solution was sterilised by autoclaving.

Solution II

0.2 M NaOH (Sigma)

1.0% (w/v) SDS (sodium dodecyl sulphate) (BDH)

This solution was made up fresh prior to use.

Solution III

60 ml 5 M potassium acetate (BDH)

11.5 ml glacial acetic acid (BDH)

28.5 ml distilled water (DW)

TE buffer (pH 8.0)

10 mM Tris-HCl (pH 8.0)

1 mM EDTA (pH 8.0)

2.7.1 Preparation of competent JM105 cells

TFN buffer

10 mM Tris-HCl, pH 8.8

50 mM $CaCl_2$ (BDH)

Volume made to 500 ml with distilled and deionised water. The buffer was sterilised by autoclaving.

JM105 bacterial strain (Pharmacia)

*sup*E, *end*A, *sbc*B15, *hsd*R4, *rps*L, *thi*, Δ(*lac-pro*AB), F' [*tra*D36, *pro*AB$^+$,*laq*I^q, *lac*ZΔM15]

10 µl of stock JM105 *E. coli* cells, stored at -70°C, were aseptically plated onto an Agar plate and grown overnight at 37°C. A single colony from the plate was then

inoculated into 10 ml of sterile LB and incubated at 37°C overnight. 500 µl of the overnight culture was aseptically inoculated into 400 ml of sterile LB in a 1 litre conical flask and incubated at 37°C on a orbital shaker (150 rpm). 1 ml aliquots of the cells were removed at intervals and measured in a spectrometer at 600 nm until the optical density (OD) of the culture was equal to 0.2 absorbance units (1 cm path length) (approximately 4½ hours incubation). The cells were harvested by centrifugation for 5 minutes at 5,000*g* in sterile 250 ml bottles (Nalgene) in an IEC PR700 centrifuge. The bacterial cell pellet was resuspended in 200 ml of ice cold TFN buffer and incubated on ice for 20 minutes. The cells were harvested by centrifugation, as before, and the bacterial cell pellet was resuspended in 20 ml ice cold TFN. 400 µl aliquots were dispensed into sterile 1.5 ml Eppendorf tubes and snap-frozen in absolute ethanol containing dry ice. The cells were immediately stored at -70°C.

2.7.2 Transformation of JM105 competent cells with recombinant pUC 18

Plasmid clones were prepared by transforming competent *E. coli* JM105 cells followed by the mini-scale preparation procedure (Section 2.7.3). 400 µl of competent JM105 cells were thawed on ice. When the cells had just thawed, 100 µl of cells were dispensed into 1.5 ml Eppendorf tubes containing approximately 15 µl of ligation mixture (Section 2.6). The cells were incubated on ice for 20 minutes and then subjected to heat-shock at 37°C for 3 minutes. 400 µl of LB broth (pre-warmed to 37°C) was added to each Eppendorf tube and the mixture was further incubated for 45 minutes at 37°C. 100 µl aliquots of the transformed cells were plated onto LB agar plates containing 50 µg/ml of ampicillin and incubated overnight at 37°C. A mini-scale preparation of plasmid DNA was prepared from single colonies using the method described in Section 2.7.3.

2.7.3 Isolation and purification of plasmid DNA from the bacterial culture

Selected colonies of transformed cells were inoculated into 5ml of LB containing 50ug/ml ampicillin and incubated overnight at 37°C, with shaking. The selected colonies were also streaked onto ampicillin, X-gal and IPTG agar plates. 5 ml overnight bacterial cultures were centrifuged at 3 000*g* in a Beckman TJ6 benchtop centrifuge for 5 minutes. The supernatants were discarded and the bacterial pellets were resuspended in 160 µl of Solution I (Section 2.7) in a 1.5 ml Eppendorf tube. 320 µl of Solution II (Section 2.7) was added, mixed, and then 240 µl of Solution III (Section 2.7) was added and mixed again. The above steps were performed for 5 min with chilled solutions and on ice. The mixture was centrifuged for 10 minutes at 13 000g. An equal volume of phenol:chloroform was added to the aqueous phase and the solution was mixed and centrifuged for 5 minutes at 13 000g. The aqueous phase was carefully removed and transferred into a new eppendorf containing 0.6 volumes of propan-2-ol (BDH). The solution was mixed and centrifuged for 10 minutes at 13 400*g* in a microcentrifuge. The DNA pellets were washed twice with 300 µl 70% ethanol and then finally dried in air. The plasmid DNA pellets were resuspended in 200 µl of TE (pH 8.0) (Section 2.7) and 5 µl of RNAse A (from a 10 mg/ml stock) (Section 2.7) was added and incubated at 37°C for 30 minutes. The solution was extracted with phenol:chloroform (Section 2.7), precipitated with ethanol (Section 2.7), washed with 70% ethanol and dried in air. The final pellet was resuspended in 20 µl of SDW.

2.7.4 Analysis of miniprep-purified recombinant plasmid DNA by restriction enzyme digestion

5X Medium salt Klenow buffer (MSK)

6.6 mM Tris-HCl pH 7.4

6.6mM $MgCl_2$

50 mM NaCl

1 mM DTT

The recombinant plasmids were analysed by digestion with restriction endonucleases. 4 µl of miniprep purified plasmid DNA was incubated with 2.5 U of *Eco* RI and 2.5 U of *Hind* III restriction endonucleases using 1 X MSK buffer for 2 hours. The digests were analysed by agarose gel electrophoresis on a 1.6% w/v agarose gel to confirm the presence of DNA inserts.

2.8 SEQUENCING OF LTR OR V3 RECOMBINANT PLASMIDS

The sequence of recombinants was determined using the Sequenase Version 2.0 T7 DNA polymerase enzyme (USB) and universal reverse and forward primers. The sequencing was performed using Sequenase Version 2.0 T7 DNA polymerase sequencing protocol (USB) as described below.

2.8.1 Denaturation of double-stranded DNA templates

5µg of the recombinant plasmid (up to 20µl) was denatured with 5µl of 1 M NaOH, 1 mM EDTA (pH 8.0) and the final volume adjusted to 25µl of SDW. The mixture was incubated at room temperature for 10 minutes. Meanwhile a CL-6B Sepharose column was prepared, this column is used to purify the DNA from the salts used for

denaturation that would interfer with the sequencing reactions. CL-6B Sepharose matrix (Sigma) was equilibrated with an equal volume of TE (pH 8.0). The CL-6B Sepharose slurry was used to fill a 0.5 ml Eppendorf tube containing a small quantity of glass beads (425-600 µm diameter) (Sigma). A small hole was pierced in the bottom of the tube with a syringe needle and the 0.5 ml Eppendorf tube was placed into a 1.5 ml Eppendorf tube, also pierced at the bottom with a syringe needle (Figure 2.1). The Eppendorf tubes were placed in a larger 15 ml centrifuge tube and centrifuged at 400*g* for 5 minutes in a Beckman TJ-6 benchtop centrifuge until the matrix was dry. The large Eppendorf was replaced by an intact sterile 1.5 ml Eppendorf. The 0.5 ml column was then ready for use. The denatured DNA solution was carefully placed on the top of the matrix and centrifuged at 400*g* for 5 minutes until all of the solution had passed through the column and was collected in the 1.5 ml Eppendorf.

2.8.2 Annealing of single stranded DNA templates to the sequencing primer

Annealing buffer

100 mM Tris-HCl pH 8.5

100 mM $MgCl_2$

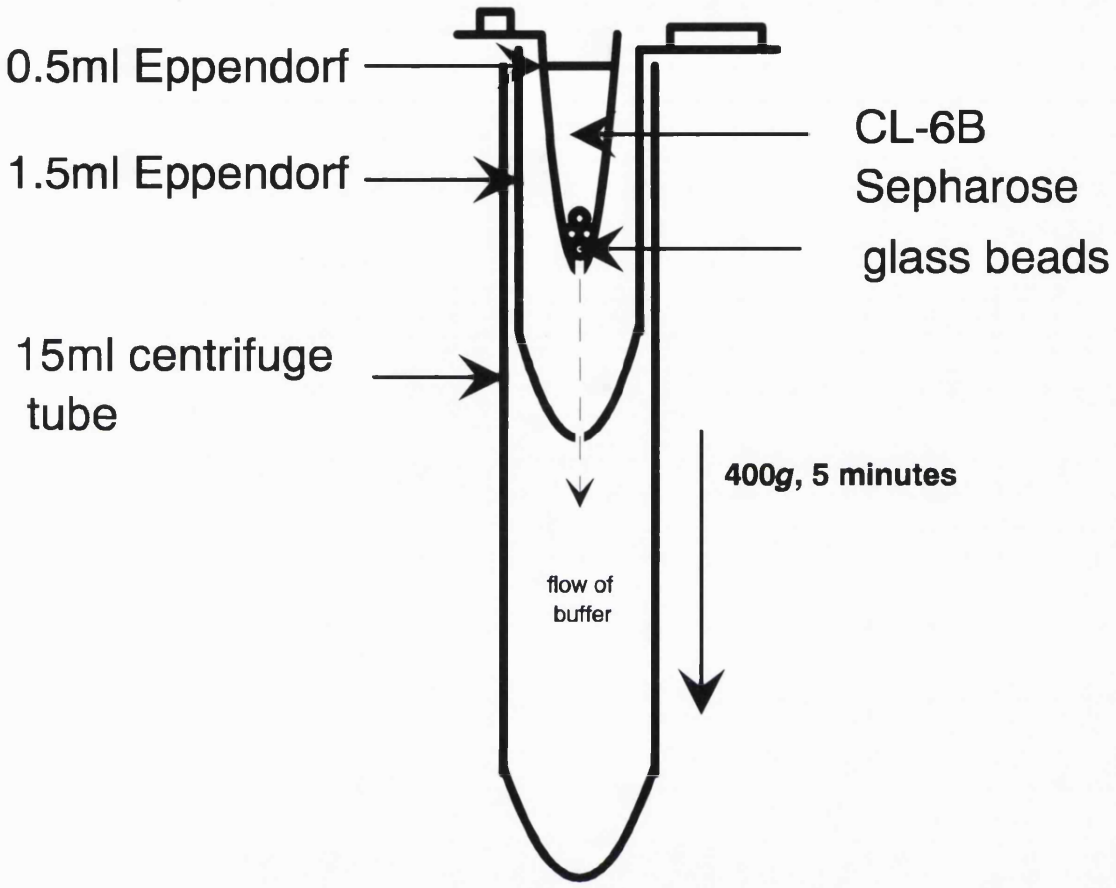


Figure 2.1: Schematic representation of the CL-6B Sepharose column used to recover a denatured DNA template.

The solution from section 2.8.1 (approximately 20 µl) was divided into two 10 µl aliquots. One aliquot was used to sequence with the (+) sense (forward) primer and the other aliquot was used to sequence with the (-) sense (reverse) primer. Primers were diluted with SDW to a concentration of 50 ng/µl. 1 µl of primer and 1 µl of annealing buffer were added to the 10 µl aliquot of denatured template DNA. The mixture was incubated at 37°C for 20 minutes and then chilled on ice.

2.8.3 Labelling of the complementary DNA strand

To each 12 µl annealed DNA template and primer mixture were added the following components in the order:

1 µl	100 mM DTT*
2 µl	1/5 dilution of Labelling Nucleotide* Mix in SDW
185 kBq	[α-^{35}S] dATP (Amersham)
2 µl	1:8 dilution of Sequenase enzyme version 2*
	(12 U/µl in Enzyme Dilution Buffer)*

*These components were supplied in the Sequenase Version 2.0 sequencing kit (United States Biochemical, USB).

The mixture was vortexed and incubated at room temperature for 2 minutes.

Labelling nucleotide mix (USB Sequenase version 2 kit)

1.5 µM each of dCTP, dGTP and dTTP

Enzyme dilution buffer (USB Sequenase version 2 kit)

10 mM Tris-HCl pH 7.5

5 mM DTT

0.5 mg/ml bovine serum albumin (BSA).

2.8.4 Termination of the labelling reaction

Stop Solution

95% (w/v) formamide (BDH)

20 mM EDTA (pH 8.0)

0.05% (w/v) bromophenol blue (BDH)

0.05% (w/v) xylene cyanol (FF) (BDH)

Four 0.5 ml Eppendorf tubes were labelled "G", "A", "T" and "C", and 2.5 µl of ddGTP (dideoxyGTP), ddATP, ddTTP and ddCTP (80µM each) were dispensed into the respective tubes. The tubes were prewarmed in a Hybaid Combi Thermal Reactor TR2 to 37°C for 1 minute and 3.5 µl of the labelling reaction described in section 2.8.3 was dispensed into each tube. The reaction was incubated at 37°C for 4 minutes and finally stopped by the addition of 4 µl of Stop Solution followed by incubation at 85°C for 5 minutes.

2.8.5 Analysis of DNA sequence by polyacrylamide gel electrophoresis

The following components (6% acrylamide mix) were dissolved on a heater/stirrer for at least 1 hour at 50°C. The container was covered to prevent evaporation.

37.5 g	urea - molecular biology grade (Sigma)
11.25 ml	40% (w/v) acrylamide (BDH) solution (19:1 acrylamide:bis-acrylamide)
7.5 ml	filtered 10X TBE (section 5.2.2.2)
26.25 ml	distilled and deionised water

After heating, the mixture was chilled on ice for 20 minutes and 450 µl of 10% (w/v) ammonium persulphate (APS) (Sigma) and 30 µl of NNN'N'-Tetramethylethylenediamine (TEMED) (BDH) were added. The acrylamide mixture was poured between two 33 x 43 cm (5 mm thick) glass plates (BRL) (both cleaned with distilled water and then 70% ethanol) one (the shorter glass plate) treated with dichlorodimethylsilane (Sigma) separated by a 0.4 mm plastic spacer (BRL). Care was taken to prevent air bubbles becoming trapped between the plates. The gel was left to set at room temperature for 30 minutes. The top of the gel was washed with 1X TBE buffer to remove any unpolymerised acrylamide and the gel was pre-run by electrophoresis at 50 W constant power in 1X TBE buffer for 30 minutes.

Approximately 5 µl of the contents of tubes labelled "G", "A", "T" and "C" were loaded into adjacent wells at the top of the gel and were subjected to electrophoresis at a constant power of 75 W until the bromophenol blue dye was approximately 3 cm from the bottom of the gel. The remaining 5 µl of solution in each tube was loaded onto a different set of four adjacent wells at the top of the gel and were subjected to electrophoresis until the bromophenol blue of this loading was 3 cm from the bottom of the gel. The sequencing plates were removed from the sequencing apparatus and the plates carefully separated so that the gel remained attached to the non-siliconised glass plate. This glass plate and the attached gel were soaked in 10% (v/v) glacial acetic acid (BDH) for 30 minutes, gently agitating the solution at regular intervals. This action both fixes the DNA within the acrylamide gel matrix and removes the urea. The gel was manually transferred onto Whatman No.3 cartridge paper and vacuum dried at 80°C for approximately 2½ hours using an Atto gel dryer and Aquavac. The dried gel was exposed to X-ray film (Hyperfilm:

Amersham plc) in a -70°C freezer in a film cassette until the bands of DNA could be observed after processing the film. The DNA sequence was read from the film over a light box.

2.9 DNA MOBILITY SHIFT ASSAYS

2.9.1 Preparation of Jurkat T cells extract

The nuclear extract preparations used in these experiments were kindly provided by Dr G. Lang from the Chester Beatty Institute (Orchard et al., 1990).

2.9.2 PAGE gel for band-shift assay

30% Bisacrylamide (4% final)	6.7ml
5 X TBE	2.5ml
Glycerol (4%)	2 ml

The gel mixture was adjusted to 50ml with DW, 250 µl of 10% APS, and 250 µl of TEMED were added to the gel mixture before pouring between glass plates to catalyze polymerisation. Once set, the gel was pre-warmed at 150 V for 30 min.

2.9.3 Labelling of Site B probe

The oligonucleotide probe used in these experiments was 30bp in length and was based on the LTR of HIV-1 strain HXB2:

5'-TCGACAGGGGTCAGATATCCACTGACCTTC

1pmol of HXB2 Site B DNA was incubated for 30 min at 37°C with dATP, dTTP, dGTT, and ^{32}PdCTP (10 mM of each); 15 U of RT; 3µl of 5X RT buffer, and adjusted to 30µl with SDW.

5X RT buffer

1mg/ml BSA

500mM Tris pH 7.6

600mM KCl

100mM MgCl

The radiolabeled probe was then purified by 5 minutes centrifugation at 400g through a 1ml Sephadex G50 column to remove unincorporated dCTP.

2.9.4 Competitive binding assays

Binding reactions were performed on ice for 20 minutes with 1µl of Jurkat-T-cell nuclear extract (6-8µg protein), 5 fmol of radiolabelled oligonucleotide probe (HXB2 site B) in 10 mM Tris HCL pH 7.8, 50mM KCl, 2mM $MgCl_2$, 0.5mM DTT, 4% glycerol, and 5µg of double stranded poly dI:dC (Pharmacia 27-7880) in a final volume of 20µl. It is important to add the nuclear extract last to the binding mixture. After incubation, the binding reaction was resolved on a 4% polyacylamide gel (see Section 2.9.2) at 4°C. Following electrophoresis, the gel was dried and autoradiographed. The oligonucleotides used were a 30bp site B probe based on the LTR of HIV-1 strain HXB2:

5'-TCGACAGGGGTCAGATATCCACTGACCTTC, and a major site B natural variant present in patients quasispecies (see Chapter 5):

5'-TCGACCAGGGCCAGGGATTAGATACCCATTGACCTTTGGATGCC.

Competition asays were performed by adding increasing (x5, x10, x20, up to x2 000 fold the amount of radioactive probe) unlabelled site B (HXB2) or site B variant in the binding reactions

2.10 COMPUTER ANALYSIS OF SEQUENCE DATA

2.10.1 Alignment of sequence data

The DNA sequences of natural LTR and V3 region variants were manually inputted via an Apple Mac computer linked to a VAX VMS running the appropriate software for analysis (Oxford University Molecular Biology Computing Unit). Manipulation of nucleotide sequences was performed using NIP (nucleotide interpretation program) available under the STADEN suite of programs. Translation of nucleotide sequences to the corresponding amino acids was performed using either NIP or TRANSLATE available within the GCG suite of programs (Sequence Analysis Software Package (Version 7) by Genetic Computers Inc. from the University of Wisconsin's Department of Genetics) (Devereux et al., 1984). Multiple sequence alignments (MSA) were produced using PILEUP, and displayed using PRETTY (both available in GCG). For phylogenetic analysis the MSA were generated using CLUSTAL V. The input file for this program is a PIR format file containing all the sequences to be aligned. The PIR was generated using TOPIR in the GCG package. Clustal alignment was created using the default settings except for the output file which was generated as a PHYLIP compatible file (Figure 2.2).

2.10.2 Median genetic divergence of quasispecies

The PHYLIP package (Version 3.5) was used in all the subsequent phylogenetic analysis (Felsenstein, 1988). The genetic distance between each variant of a quasispecies was calculated using a Kimura-2 parameter method (available within DNADIST) (see Figure 2.2). The median divergence was calculated from all the pair-wise genetic distances of variants from the quasispecies.

2.10.3 Construction of phylogenetic trees

Maximum parsimony trees were generated using DNAPARS. The robustness of the trees was tested by bootstrap resampling of the multiple sequence alignment using SEQBOOT (100 datasets) and consensus trees generated using CONSENSE. For a limited number of datasets, trees were also constructed using FITCH which uses the output from the distance matrix. All the trees presented in this thesis were generated using DNAPARS because it has been shown to be of equal performance to Maximum likelihood and Neighbour-joining but less computationally intensive (Hillis et al., 1994).

2.11 STATISTICAL ANALYSIS

Analysis of the significance of differences in genetic distances calculated from the output of the PHYLIP DNADIST (Section 2.10.2) was performed using the Mann-Whitney U-test (Altman 1993).

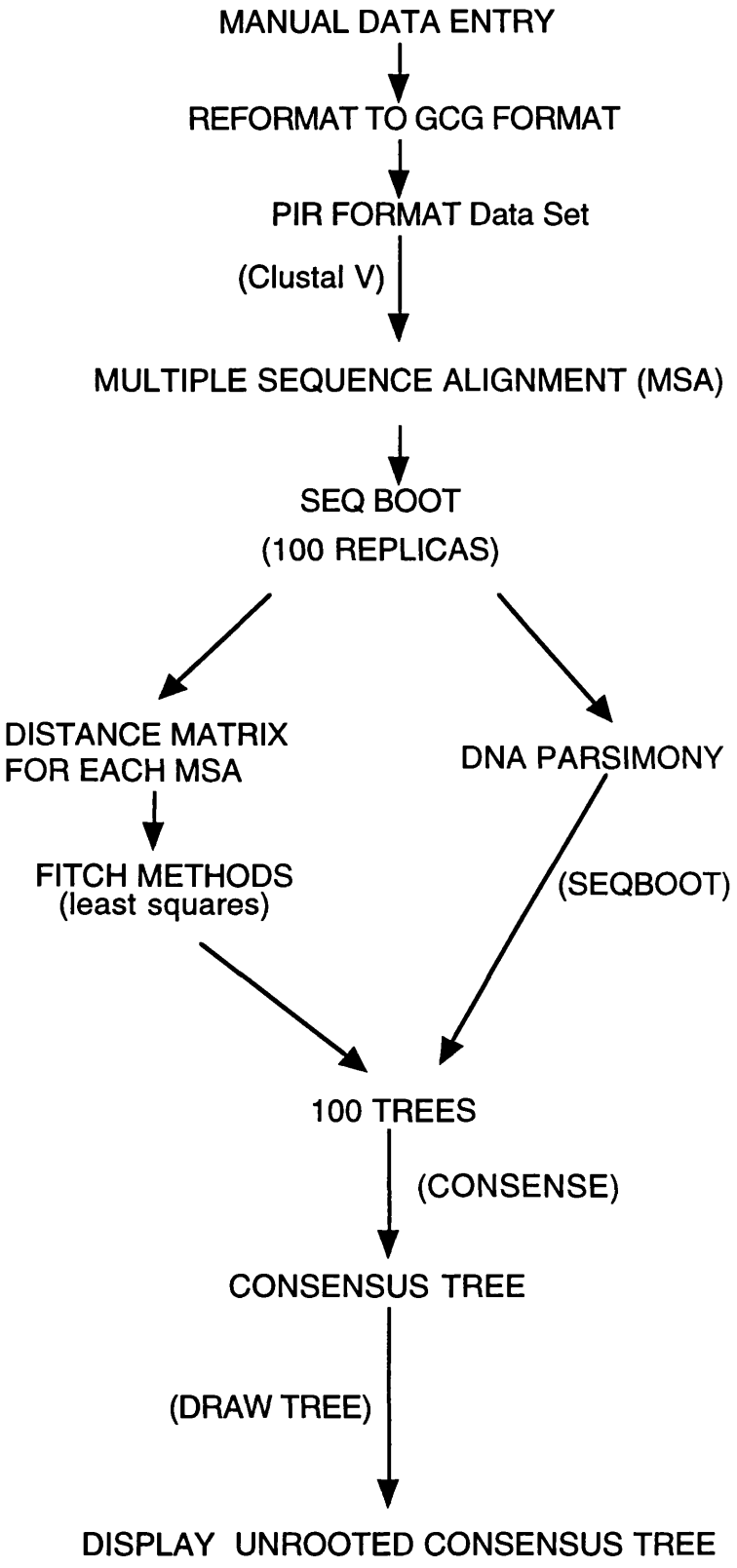


Figure 2.2: Schematic representation of the computing procedures used for sequence data analysis.

CHAPTER 3

OPTIMISATION OF HIV-1 LTR AND V3 REGION SPECIFIC PCRs, VALIDATION OF SEQUENCING STRATEGY, AND EVALUATION OF ULTRAVIOLET LIGHT (UV) IRRADIATION TO ELIMINATE PCR CONTAMINATION

3.1 INTRODUCTION

The high sensitivity of the polymerase chain reaction (PCR), an amplification method that can potentially detect 1 molecule of target DNA, means that even very low levels of contamination with exogenous, either plasmid-derived, or previously generated amplicons will result in a false positive signal. This aspect is a major limitation in the use of PCR as a routine diagnostic method (Bell, 1989; Clewley, 1989). This ease of contamination can also be a problem in the use of PCR in a research context, for example, in quantitative PCR assays or the study of pathogen prevalence in a patient group. Indeed, certain PCR-based studies, for example the reported prevalence of human papillomavirus type 16 in women with no cervical abnormalities have subsequently been invalidated due to contamination (Tidy et al., 1989). A method of overcoming PCR contamination using UV light has been reported (Sarkar and Sommer, 1990). The major effect of UV light on DNA is to induce cyclobutane pyrimidine dimers which can inhibit DNA replication by DNA polymerase (Moore et al., 1981). Subsequent to the original report by Sarker and Sommer (1990), it has been proposed (Cimino et al., 1990) that UV irradiation is only useful as a PCR decontamination procedure when the target sequence is relatively large (approximately 500 bp). This limitation of target DNA size was investigated by Fox and colleagues (Fox et al., 1991) who showed that UV irradiation was effective in eradicating contamination for the PCR of a 149 bp fragment corresponding to a region of the glycoprotein B gene of CMV, a 375 bp fragment corresponding to an early gene promoter of CMV and a 541 bp fragment of the HIV-1 LTR. In this study, the investigators showed that UV irradiation was effective when contaminant DNA was present in an amount below 100pg. However, they showed that the ultimate sensitivity of the PCR system to amplify the 375 bp fragment was reduced 1000-fold. In this chapter, the optimisation of a PCR used

to amplify a 541 bp fragment of the HIV-1 LTR and V3 region, as well as the detailed study of the effects of UV irradiation on sensitivity and decontamination efficacy of this PCR system are presented.

The second part of this chapter, aimed at evaluating the extent of mutations induced by the Taq polymerase used in all PCR assays carried out in this thesis. The fidelity of DNA polymerases used in PCR are influenced by many factors, such as, pH, concentrations of deoxynucleoside triphosphates, and magnesium ion (Keohavong et al., 1993; Cariello et al., 1991). The Taq polymerase used in the present study lacks a 3' to 5' exonuclease activity or proofreading activity. Several investigators reported estimates of the Taq error rates under the optimised amplification conditions for various template DNA amplification. The error rates reported in error/base pair incorporated were in the 10^5 range, 7.2×10^{-5} (Ling et al., 1991), 2.0×10^{-5} (Lundberg et al., 1991), and 8.9×10^{-5} (Cariello et al., 1991).

Therefore, it was important to estimate the Taq error rate under the conditions that permitted efficient amplification of the 2 HIV-1 loci (LTR and V3 region) amplified in this study. In order to do so, the DNA sequence of 15 LTR and 15 V3 region clones respectively derived from a single known sequence of LTR and V3 region was determined using standard methods (see Sections 2.5, 2.6, 2.7, and 2.8). The Taq error rate under the optimised conditions of amplification used in this thesis was calculated as follows:

DNA template length x Number of PCR cycle x number of clones sequenced and read x number of mutations generated from the original template.

3.2 LTR PCR OPTIMISATION

The LTR PCR was carried out on a Hybaid thermal reactor (model HBTR1) as described originally (Saiki et al., 1988) except that a final magnesium concentration of 1.5 mM was used in the Taq buffer. The primer sequences used were:

LTR A: 5'-CACACAAGGCTACTTCCCTGA

LTR B: 5'-GATCTCTAGTTACCAGAGTCAC

The PCR produced a 541bp fragment according to the HIV-1 HXB2 prototype.

3.2.1 Magnesium titration

The PCR was carried out with 200ng of each primer, 500 copies of cloned template and increasing magnesium concentrations ranging from 1 to 5 mM final concentration, in 1mM increments. The best amplification of target was obtained between 1.5 and 3 mM. Thus, the same procedure as above was repeated with magnesium concentrations ranging from 1.5 to 3 mM final concentration in 0.5 mM increments. The most specific amplification of target was obtained with a 1.5 mM final concentration of magnesium (Figure 3.1). This concentration was subsequently used in all the PCR reactions.

3.2.2 Primer titration

The PCR was carried out at the optimum magnesium concentration of 1.5 mM final with increasing amount of each primer, ranging from 50 to 230 ng and 500 copies of cloned template. At this magnesium concentration, the least non-specific bands were obtained with 125 ng of each primer (Figure 3.2, lane 4).

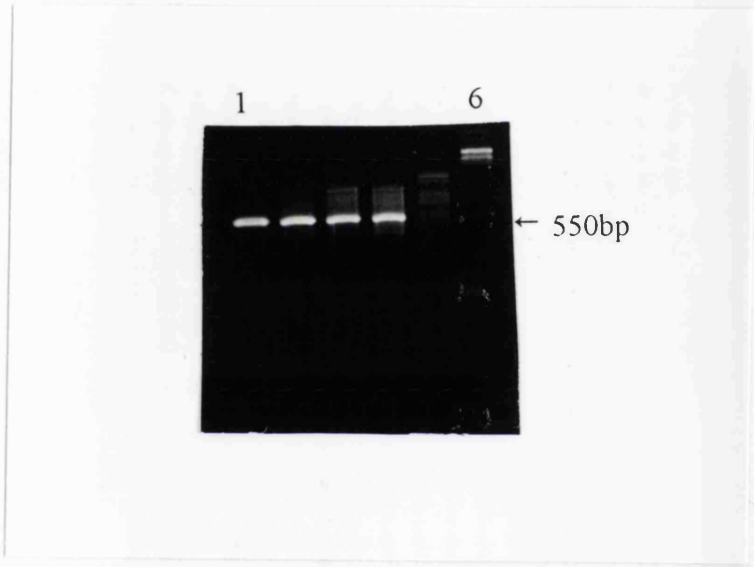


Figure 3.1: Magnesium titration of the LTR PCR. Lanes 1 to 4 represent the LTR PCR at 1.5, 2, 2.5, and 3mM respectively. Lane 5 and 6 contains DNA markers, PBR322 A*va* II and Lambda H*ind* III respectively.

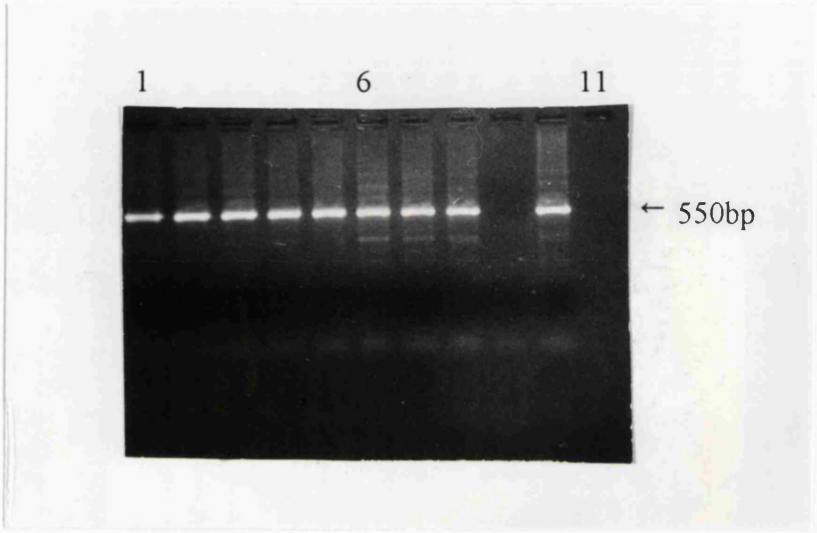


Figure 3.2: Primer titration. Lanes 1 to 8 represent LTR amplicons resulting from PCRs carried out with 50, 75 100, 125, 150, 175, 200, and 230 ng of primers respectively. Lane 9 is a negative control and lane 10 is a positive control (10^4 copies of cloned target).

3.2.3 Template titration: PCR sensitivity

The sensitivity of the HIV-1 LTR PCR under optimum magnesium and primer concentrations was determined by amplifying decreasing amounts of target DNA ranging from $5x10^7$ to 0.2 copies of a plasmid containing the HIV-1 LTR . Under these optimised conditions, the PCR system was able to detect 2 copies (lane 7) of DNA template (Figure 3.3). This represents an optimal sensitivity that would slightly decrease in the presence of human cellular DNA, the specificity of the PCR is described in the section that follows.

3.2.4 Primer concentration and sensitivity with clinical material

To ensure that the low primer concentration does not affect the sensitivity of the PCR system when using clinical samples, parallel PCRs were performed with 125ng and 250ng of primers and the same serial dilution of patient 3 lymph node DNA extract using 1/10 serial dilution of 1µg of DNA. As shown in figure 3.4, increasing the amount of primers from 125ng to 250ng in the PCR did not increase the sensitivity of detection on clinical material but resulted in an increase in the number and intensity of non-specific bands.

3.2.5 V3 region amplification

The V3 region PCR was optimised using the same strategy as that used for the LTR PCR and is described in Section 2.4.1. The optimum conditions were 150ng of each primer, 4mM magnesium chloride final concentration, and a 60°C annealing temperature (see table 2.2).

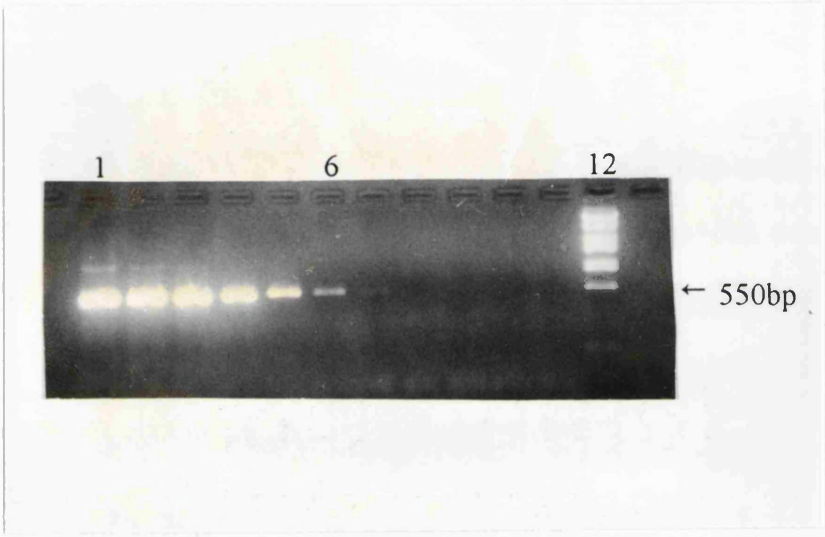


Figure 3.3: Template titration. Lanes 1 to 8 represents LTR products resulting from PCR with $5X10^7$, $5x10^6$, $5x10^5$, 500, 50, 5, 2, and 0.2 copies of template DNA respectively. Lanes 9 and 10 are PCR negative controls. Lane 12 contains 1µg of Lambda H*ind* III marker.

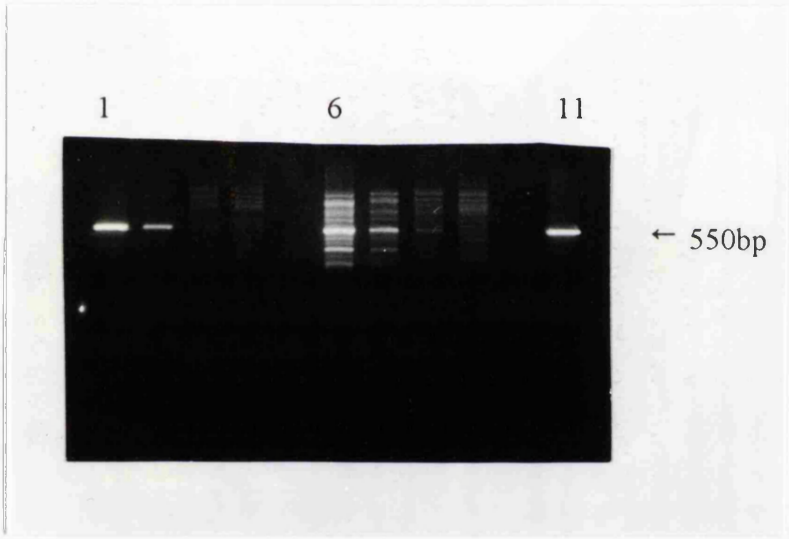


Figure 3.4: PCR sensitivity with 125ng and 250ng of primers. Lanes 1 to 4 represent the LTR PCR of 1 µg, 100ng, 10ng and 1ng of patient 3 lymph node DNA respectively using 125ng of primers. Lanes 6 to 9 represent the LTR PCR of 1 µg, 100ng, 10ng and 1ng of patient 3 lymph node DNA (same source as lanes 1 to 4) respectively using 250ng of primers. Lanes 5 and 10 are negative controls. Lane 11 is a positive control.

3.3 UV-IRRADIATION FOR ELIMINATING CONTAMINATION

Irradiation was carried out using two wavelengths of UV light simultaneously, 302 and 254nm, delivered by a UVP transilluminator (model TM15; Genetic Research Instrumentation, Cambridge) and a UVP hand-held lamp (model UVGL-50; Genetic Research Instrumentation) respectively. All the reagents for the PCR, including primers and Taq polymerase, but excluding target DNA, were dispensed into 0.5 ml polypropylene eppendorf tubes (Alpha Laboratories) resulting in a final volume of 90µl per tube. Duplicate tubes were exposed to UV irradiation for 5 minutes by placing them directly onto the screen of the transilluminator, with the hand-held lamp positioned above the tubes, "sandwiching" them between the two UV sources, thus maximising the dose delivered to the PCR mixture. After UV exposure template DNA was added and the tubes placed in the thermal reactor. The effect of UV irradiation on the PCR sensitivity was investigated by comparing the detection threshold of the PCR system when the PCR reagents are subjected to irradiation before addition of target. The efficacy of UV exposure to decontaminate the PCR was investigated by carrying out the same procedure as above but adding a range of plasmid DNA template prior to UV exposure. Target DNA ranged from 10^6 to 5 copies of recombinant pUC18 plasmid with the HIV-1 LTR fragment inserted in the polycloning site. It is worth noting that plasmid DNA is mainly in a supercoiled form, and therefore, amplicons or proviral sequences from clinical specimens may be easier to cross-link resulting in more efficient eradication of such contaminants.

3.3.1 Effect of UV irradiation on PCR sensitivity

As shown in figure 3.5, the intrinsic sensitivity of the PCR system was not affected by UV irradiation, still detecting 2 copies of DNA template. However, the overall efficiency of the reaction must be affected since the intensity of the amplimers is

decreased when the PCR reagents were subjected to UV irradiation before addition of target (Figure 3.5).

3.3.2 Efficacy of UV irradiation to eradicate PCR contamination from pre-added template

To mimic contamination, template DNA (same source as that used in section 3.2.1 (figure 3.5)) was added prior to UV irradiation, then, subjected to PCR under the optimised conditions (Section 3.2). As shown in figure 3.6, UV irradiation resulted in the successful eradication of at least 500 copies of template DNA. However, 1000 copies or greater were still detected by the PCR despite 5 minutes UV irradiation.

3.4 VALIDATION OF SEQUENCING DATA BY SEQUENCING MULTIPLE LTR AND V3 REGION CLONES DERIVED FROM A SINGLE ORIGINAL SEQUENCE

3.4.1 PCR and cloning of LTR and V3 variants derived from clinical specimens

PCR assays were carried out using 0.5 or 1 µg of DNA extracted from the clinical samples investigated in this thesis (see Chapter 2, Sections 2.2.1, 2.3, and 2.4). The amplicons derived from 4 PCRs for each specimen were jointly purified and subsequently cloned in pUC18 (see Sections 2.5, 2.6, and 2.7). The recombinant plasmids with the inserts of interest were identified by restriction enzyme digestion of plasmid mini-scale preparations of single selected colonies (see Sections 2.7.3, and 2.7.4). The efficiency of ligation ranged from 66% (Figure 3.7 (a)) to 100% (Figure 3.7 (b)). The recombinant plasmids with the appropriate size inserts were subsequently sequenced using the method described in section 2.8.

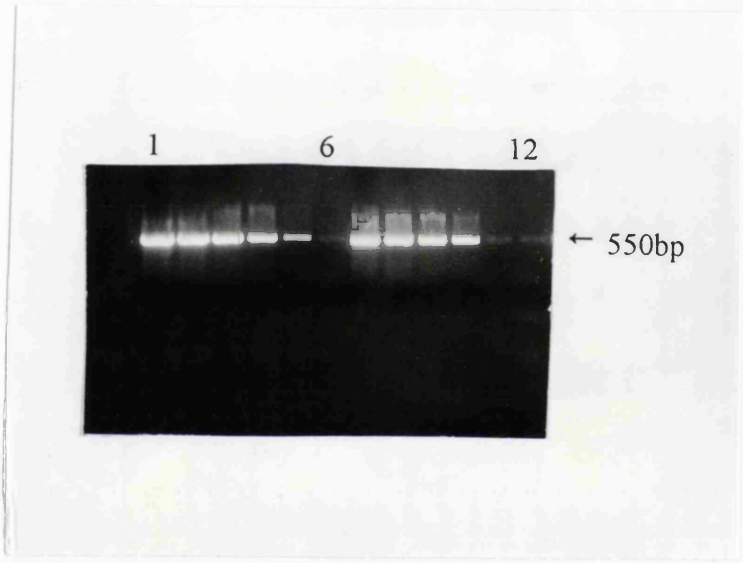


Figure 3.5: Effect of UV irradiation on PCR sensitivity. Lanes 1 to 6 represent PCRs with 10^6, 10^4, 10^3, 500, 50, and 2 copies of cloned template without UV irradiation. Lanes 7 to 12 represent PCRs with 10^6, 10^4, 10^3, 500, 50, and 2 copies of cloned template with UV irradiation of PCR mixture prior to addition of template.

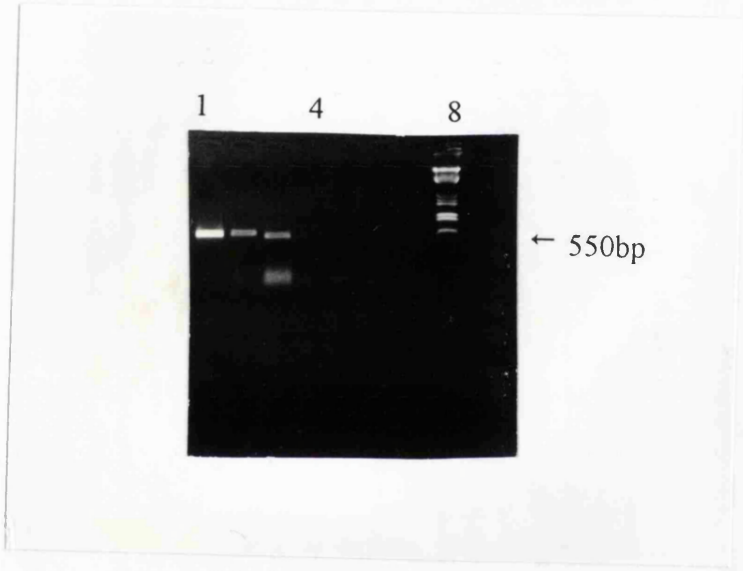


Figure 3.6: Efficacy of UV irradiation to eradicate PCR contamination. Lanes 1 to 6 represent PCRs of 10^6, 10^4, 10^3, 500, 50, and 2 copies of cloned template with 5 min UV irradiation of the PCR mixtures containing the target DNA. Lane 7 is a negative control. Lane 8 contains a DNA marker, 700ng of Lamda H*ind* III.

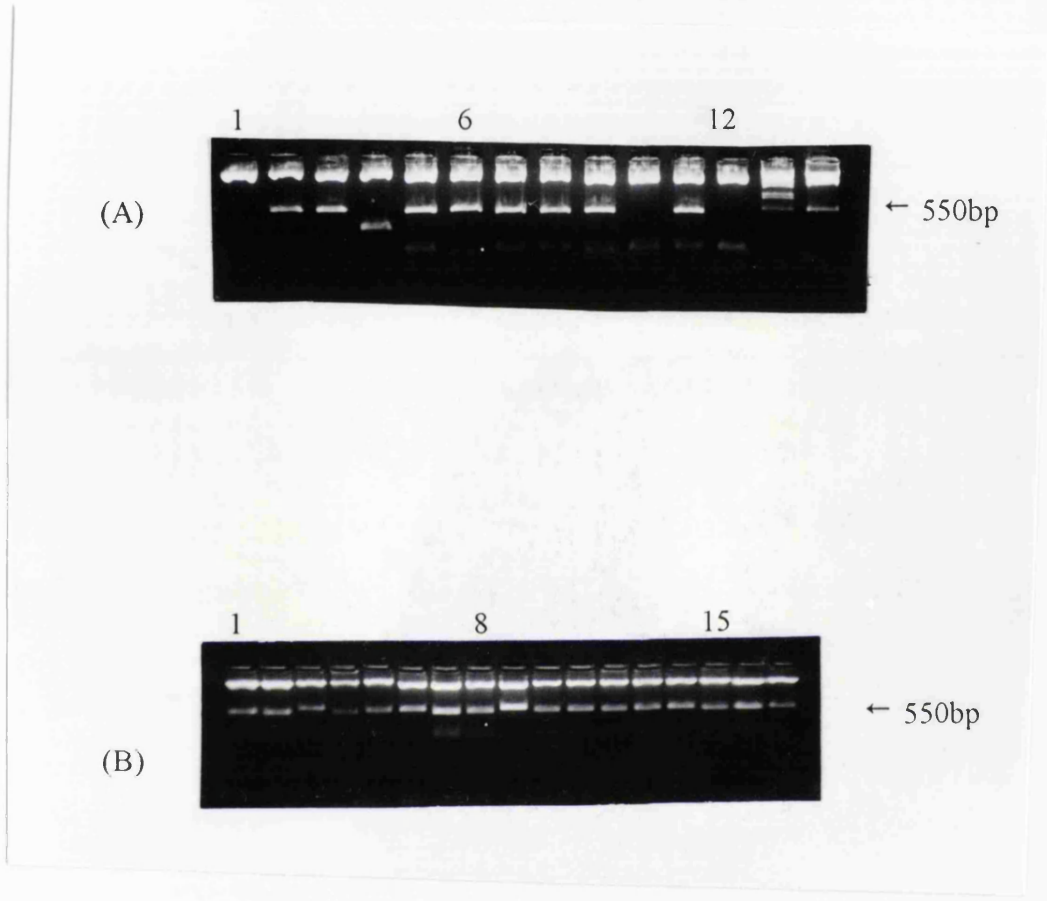


Figure 3.7: Recombinant plasmid mini-preparations restricted with E*co* RI and H*ind* III. (a) Lanes 1 to 12 contain restricted plasmid mini-preparations derived from 12 distinct colonies transformed with the ligation product of PBMC LTR amplicon of patient 1 and S*ma* I digested pUC 18. Lane 13 contains the DNA marker Lambda H*ind* III. (b) Lanes 1 to 17 contain restricted plasmid minipreparations derived from 17 distinct colonies transformed with the ligation product of PBMC LTR amplicon of patient 2 and pUC 18.

3.4.2 Evaluation of *Taq* error rates for the LTR and V3 region optimised PCR assays

Considering the relatively high error rate of *Taq* polymerase, in the order of 10^{-5} error/base pair incorporated, it is crucial to verify the validity of the DNA sequence data derived from PCR ampification. In order to do so, the DNA sequence of 15 LTR and 15 V3 region clones respectively derived from a single known sequence of LTR (clone 1 of patient 1 PBMC derived LTR) and V3 region (clone 1 of patient 0 V3 region) was determined using standard methods (see Chapter 2, Section 2.8). For the LTR and V3 plasmids, 10^3 copies of clone were used to re-amplify the sequences. The sequence of each LTR and V3 region clone was identical to the sequence of the original PCR template. This experiment allowed the calculation of the Taq error rate under the conditions used for both the LTR and V3 region amplification. The Taq error rate under the LTR amplification conditions, and using 550bp as the average lentgh of the LTR amplicon is:

< 550bp x 35 cycles x 15 clones = 2.8×10^{-5} errors/bp incorporated.

The taq error rate under the V3 region amplication conditions, and using 361bp as the average lentgh of the amplicon is:

< 361bp x 35 cycles x 15 clones = 1.8×10^{-5} errors/bp incorporated.

It was therefore concluded that any mutations observed in the multiple sequence data derived from clinical specimens in the present study are true mutations (not *Taq* induced) and represent natural variant genomes present in the specimen.

3.5 DISCUSSION

This chapter summarizes the optimisation strategies of PCR assays. The optimised conditions for the LTR PCR were 1.5 mM magnesium, 125 ng of primer, and 60°C annealing temperature. For the V3 region PCR, the optimised conditions were 4 mM magnesium, 150 ng of primer, and 60°C annealing temperature. I have used these conditions throughout this thesis.

The extreme sensitivity of the PCR is a major advantage when considering its feasibility as a diagnostic assay or as a molecular biology tool for the amplification of infective agent nucleic acid present at very low levels in the host. However, this high sensitivity, means that very low levels of contamination with exogenous DNA will result in erroneous data. The second part of this chapter investigated the efficacy of UV irradiation in preventing PCR contamination and the effect this procedure might have on the sensitivity of the HIV-1 LTR-PCR system.The results indeed showed that UV irradiation of the PCR mixture with pre-added template can eradicate contamination of at least 500 copies of exogenous DNA template without recourse to intricate time consuming manipulations. The results obtained by Fox and colleagues (1991) showed that UV-irradiation can adversey affect the sensitivity of some PCR systems. The results obtained in this chapter showed that the HIV-1 LTR PCR which was sensitive to 2 copies of template DNA was not affected by the UV irradiation procedure. This was not the case for a 347 bp early gene promoter of CMV PCR system where the sensitivity was reduced by a 1000-fold following irradiation (Fox et al., 1991). The reasons for the different extents of decreased sensitivity caused by UV irradiation are unclear. A possible explanation is that variations in the primer DNA sequences, ie, the number of adjacent pyrimidines with potential to form pyrimidine dimers, or primer concentrations in the different PCR

systems result in different levels of primer DNA damage by irradiation which differentially affect primer/target DNA binding, thus reducing the efficiency of certain PCRs. The data support the view that UV irradiation of PCR components is able to eradicate exogenous contamination to an extent (500 copies in the case of the LTR PCR). However, the effects of UV irradiation on the sensitivity of the PCR must be determined empirically for each PCR system to ensure that this aspect is not compromised. Furthermore, UV irradiation should be seen as an additional precaution rather than a replacement for careful laboratory practice. Indeed, the most efficient method to avoid PCR contamination is the physical separation of the PCR components and the target DNA. This is achieved by undertaking the PCR method in a multi-room setting. This setting consists of a DNA-free room, where all the PCR reagents (primers, Taq buffer, magnesium solution, dNTPs, Taq polymerase, sterile distilled water and mineral oil) are kept. This clean PCR room has allocated pipettes, aerosol-resistant tips, racks and gloves. Under no circumstances, should any material that has been in contact with DNA (i.e. coming from another research laboratory) be brought into the clean PCR room. Finally, surgical gowns are available in this room and should be worn at all times. The closed PCR tubes are then taken to the PCR set-up room, where the target DNA (clinical material) is added in a class 2 safety cabinet. This PCR set-up room is free of plasmid DNA or any positive control at a high concentration. The addition of template is carried out with allocated pipettes and aerosol resistant tips. A negative control is included in the clean PCR room to check for possible contamination of any of the reagents. The other negative controls are formulated in the class 2 cabinet of the PCR set-up room; one negative control is introduced for every 5 analyses. The positive control can be added in any of the molecular research laboratories. The PCRs are then subjected to thermal cycling in a third room containing the

thermocyclers. Throughout this thesis, I have used this 3-room format when performing my PCRs.

Considering the relatively high error rate of Taq polymerase, in the order of 10^{-5} error/nucleotide incorporated, the final part of this chapter aimed at validating the authenticity of the DNA sequence data obtained from clones of PCR amplicons. In order to address this, the DNA sequence of 15 LTR and 15 V3 region clones respectively derived from a single known previously sequenced HIV-1 LTR clone (clone 1 of patient 1 PBMC derived LTR) and a V3 region clone (clone 1 of patient 0 V3 region) was determined. The sequence of each LTR and V3 region clone was identical to the sequence of the original PCR template. This experiment allowed the calculation of the Taq error rate under the conditions used for both the LTR and V3 region amplification. The Taq error rate under the LTR amplification conditions, and using 550bp as the average length of the LTR amplicon, and the V3 region PCR conditions with 361 bp as a average length of the amplicon were:
$<2.8 \times 10^{-5}$ errors/bp incorporated, and $<1.8 \times 10^{-5}$ errors/bp incorporated respectively. The above error rate estimates for Taq are in good accord with the ones previously reported, 7.2×10^{-5} (Ling et al., 1991), 2.0×10^{-5} (Lundberg et al., 1991), and 8.9×10^{-5} (Cariello et al., 1991).

It was therefore concluded that any mutations observed in the multiple sequence data derived from clinical specimens in the present study are true mutations (not *Taq* induced) and represent natural variant viruses present in the specimen.

CHAPTER 4

SEQUENCE ANALYSIS OF HIV-1 V3 REGION PROVIRAL BLOOD VARIANTS FROM A PATIENT UNDERGOING SEROCONVERSION AND A PATIENT AT THE LATE STAGES OF HIV DISEASE.

4.1 INTRODUCTION

4.1.1 Sequence variation of the V3 region at seroconversion

Significant genetic variation occurs within the human immunodeficiency virus (HIV) genome (see Section 1.3) (Wain-Hobson, 1993). Many studies have shown that variation occurs in both an inter- and intra-patient fashion (Delassus et al., 1991; Holmes et al., 1995; Donaldson et al., 1994a; Balfe et al., 1990; Cheynier et al., 1990)and that heterogeneity is not evenly distributed throughout the genome (Pedroza Martins et al., 1992; Simmonds et al., 1990). Thus, the hypervariable regions (V1-V5) of the *env* gene are significantly more heterogeneous than regions of the *gag* gene (Brown and Monaghan, 1988; Balfe et al., 1990; Simmonds et al., 1990; Brown, 1991) or the LTR (Delassus et al., 1991; Delassus et al., 1992b; Michael et al., 1994; McNearney et al., 1995; Ait-Khaled and Emery, 1994; Ait-Khaled et al., 1995).

The *env* gene contains the principal neutralization determinant (PND) of HIV-1 (Rusche et al., 1988; Palker et al., 1988). It consists of a loop of 35 amino acids secured by a disulphide bridge between two cysteines within the third variable region of *env* (the V3 loop). Mutations within this region have been shown to segregate with tissue tropism (Chesebro et al., 1991; Chesebro et al., 1992; Epstein et al., 1991; Westervelt et al., 1992; O'Brien et al., 1990; Zhu et al., 1993; Kuiken et al., 1992; de Jong et al., 1992a) and to be involved in the syncytium inducing phenotype (de Jong et al., 1992b; Grimaila et al., 1992). In addition, several studies have shown that despite evolution of V3 variants, the crown of the loop invariably contains the GPGR motif, although minor variants are apparent (LaRosa et al., 1990). Variation in the *env* gene is likely to be a consequence of selective pressure either through

B-cell or T-cell immune responses and yield virus strains that could evade the prevailing immune surveillance. However, variation is also constrained by the requirement to retain the biological functions associated with the V3 region.

HIV-1 Primary infection is accompanied by high levels of plasma viraemia as assessed by both p24 production and viral load measurements (Daar et al., 1991; Clark et al., 1991). At this stage of infection antibodies to HIV-1 are usually not detected. HIV specific antibodies can then be detected 3 to 6 months following infection. In some individuals an acute seroconversion illness can occur. On the basis of molecular epidemiological data (Brown, 1991), it is likely that the primary inoculum contains a multiplicity of variants which are not necessarily equally able to infect and replicate in the primary target cell: the CD4+ monocytes/macrophages in the case of sexual transmission when the submucosal barrier has to be overcome or the CD4+ lymphocytes in the case of blood to blood transmission (haemophiliacs and IVDUs). At present, relatively little data are available on the complexity of quasispecies present in the host at time of seroconversion. Since the host does not produce an immune response until after the initial viraemia, the V3 sequence within *env* should not be subject to B-cell mediated immune selective pressure, although T-cell pressure may be apparent at an early stage. In this chapter, I have determined the V3 quasispecies present in the proviral DNA from an individual (patient 0) undergoing seroconversion to HIV-1 to investigate the constraints present on the V3 region during the early stages of HIV infection; and compared it with the V3 region variants present in a patient with AIDS (see Chapter 1 section 4.1.2).

4.1.2 Sequence variation in the V3 region at late stages of HIV-1 infection

As discussed in Section 1.6, the lymphoid tissues have been shown to be a major reservoir of virus, both as virions trapped on the Fc and C3b receptors of follicular dendritic cells and as integrated provirus in CD4+ lymphocytes some of which produce virus. Indeed, several groups have reported a high viral load in lymph nodes at all stages of HIV-1 infection. In contrast, the proviral load in the peripheral blood is low at early and intermediate stages of HIV-1 infection, although a higher load is observed at late stages of disease (Pantaleo et al., 1993a; Embretson et al., 1993). One hypothetical explanation for this effect is the seeding of HIV-1 strains from the lymph node to the periphery following the extensive destruction of the lymph node histological structure and therefore its immunological competence (Frost and McLean, 1994). In contrast to the data summarised in section 4.1.1, the V3 region rapidly evolves during the clinical latency of HIV-1 disease. The increased diversity of the V3 region, has been correlated with phenotypic changes of the virus, including increased virulence (NSI/SI), broadening of tropism, and immune escape. These correlations were discussed in detail in chapter 1, sections 1.3.1, 1.3.2, and 1.3.3. The second part of this chapter presents the V3 region sequence variation of proviral DNA isolated from a lymph node biopsy and a simultaneous blood sample of a symptomatic patient (patient 1) with CDC IV disease. The V3 loop aa sequence will be analysed with respect to the well defined aa positions involved in antigenic variation and NSI or SI phenotypes (see Chapter 1 section 1.3.1). The phylogenetic analysis of the V3 region isolated from blood and lymph node of patient 1 presented in this chapter aimed at investigating the possible role of the lymph nodes as a site(s) of strain evolution and the relationship of variants present in various infected organs.

4.2 HIV-1 V3 REGION SEQUENCE VARIATION AT DIFFERENT STAGES OF INFECTION

4.2.1 Analysis of HIV-1 V3 region sequences isolated from peripheral blood of patient 0

Patient 0 presented at day 0 with diarrhoea and rash. 13 days later, after exclusion of bacterial pathogens, HIV seroconversion illness was suspected and antibodies to HIV-1 (Abbott EIA) found to be equivocal. At the same time HIV antigens (Abbott) were positive (15pg/ml). Two days later, HIV antigens became negative and antibodies to HIV-1 were strongly positive. The patient had elevated β_2m levels (3.3mg/l), a depressed CD4% (18%), and a CD4/CD8 ratio of 0.9 (normal range 0.7-3.3). A PBMC sample in preservative free heparin was obtained from the patient at day 13 after presentation. The HIV-1 V3 region from proviral DNA present in the PBMC of patient 0 has been amplified by PCR, cloned into the Sma I restriction site of pUC18 and DNA sequences from the resultant clones were determined. The DNA sequences of 15 V3 cloned amplimers were determined. The 15 V3 sequences were subjected to sequential pairwise alignment to produce the multiple alignment shown in Figure 4.1.

```
      1                                                            50
Cor5  ---------- ---------- ---------- ---------- ----------
Cor6  ---------- ---------- ---------- ---------- ----------
Cor7  ---------- ---------- ---------- ---------- ----------
Cor8  ---------- ---------- ---------- ---------- ----------
Cor2  ---------- ---------- ---------- ---------- ----------
Cor9  ---------- ---------- ---------- ---------- ----------
Cor11 ---------- ---------- ---------- ---------- ----------
Cor13 ---------- ---------- ---------- ---------- ----------
Cor3  ---------- ---------- ---------- ---------- ----------
Cor1  ---------- ---------- ---------- ---------- ----------
Cor10 ----c----- ---------- ---------- ---------- ----------
Cor14 ---------- ---------- ---------- ---------- ----------
Cor12 ---------- ---------- ---------- ---------- ----------
Cor15 ---------- ---------- -------t-- ---------- ----------
Cor4  ---------- ---------- ---------- ---------- ----------
Cons. CGCTAGGAAT TCGGCCAGTA GTATCAACTC AACTGCTGTT AAATGGCAGT

      51                                                           100
Cor5  ---------- ---------- ---------- ---------- ----------
Cor6  ---------- ---------- ---------- ---------- ----------
Cor7  ---------- ---------- ---------- ---------- ----------
Cor8  ---------- ---------- ---------- ---------- ----------
Cor2  ---------- ---------- ---------- ---------- ----------
Cor9  ---------- ---------- ---------- ---------- ----------
Cor11 ---------- ---------- ---------- ---------- ----------
Cor13 ---------- ---------- ---------- ---------- ----------
Cor3  ---------- ---------- ---------- ---------- ----------
Cor1  ---------- ---------- ---------- ---------- ----------
Cor10 ---------- ---------- ---------- ---------- ----------
Cor14 ---------- ---------- ---------- ---------- ----------
Cor12 ---------- ---------- ---------- ---------- ----------
Cor15 ---------- ---------- ---------- ---------- ----------
Cor4  ---------- ---------- ---------a ga-------- ----------
Cons. CTATCAGAAG GAGAGGTAGT AATTAGATCT AGCAATTTCA CGGACAATGC
```

Figure 4.1: Multiple sequence alignment of 15 V3 region variants isolated from patient 0. The consensus sequence derived from the alignment is also shown. Deletions are indicated as points and similarities to the consensus sequence are shown as dashes.

```
     101                                                        150
Cor5  ---------- ---------- ---------- ---------- ----------
Cor6  ---------- ---------- ---------- ---------- ----------
Cor7  ---------- ---------- ---------- ---------- ----------
Cor8  ---------- ---------- ---------- ---------- ----------
Cor2  ---------- ---------- ---------- ---------- ----------
Cor9  ---------- ---------- ---------- ---------- ----------
Cor11 ---------- ---------- ---------- ---------- ----------
Cor13 ---------- ---------- ---------- ---------- ----------
Cor3  ---------- ---------- ---------- ---------- ----------
Cor1  ---------- ---------- ---------- ---------- ----------
Cor10 ---------- ---------- ---------- ---------- ----------
Cor14 ---------- ---------- ---------- ---------- ----------
Cor12 ------t--- ---------- ---------- ---------- ----------
Cor15 ---------- ---------- ---------- ---------- ----------
Cor4  ---------- ---------- ---------- ---------- ----------
Cons. TAAAATCATA ATAGTACAGC TGAATAAATC TGTAGAAATT AATTGTACAA

     151                                                        200
Cor5  ---------- ---------- ---------- ---------- ----------
Cor6  ---------- ---------- ---------- ---------- ----------
Cor7  ---------- ---------- ---------- ---------- ----------
Cor8  ---------- ---------- ---------- ---------- ----------
Cor2  ---------- ---------- ---------- ---------- ----------
Cor9  ---------- ---------- ---------- ---------- ----------
Cor11 ---------- ---------- ---------- ---------- ----------
Cor13 ---------- ---------- ---------- ---------- ----------
Cor3  ---------- ---------- ---------- ---------- ----------
Cor1  a--------- ---------- ---------- ---------- ----------
Cor10 ---------- ---------- ---------- ---------- ----------
Cor14 ---------- ---------- ---------- ---------- ----------
Cor12 ---------- ---------- ---------- ---------- ----------
Cor15 ---------- ---------- ---------- ---------- ----------
Cor4  ---------- ---------- ---------- ---------- ----------
Cons. GACCCAACAA CAATACAAGA AAAAGTATAC ATATAGGACC AGGCAAAACA

     201                                                        250
Cor5  ---------- ---------- ---------- ---------- ----------
Cor6  ---------- ---------- ---------- ---------- ----------
Cor7  ---------- ---------- ---------- ---------- ----------
Cor8  ---------- ---------- ---------- ---------- ----------
Cor2  ---------- ---------- ---------- ---------- ----------
Cor9  ---------- ---------- ---------- ---------- ----------
Cor11 ---------- ---------- ---------- ---------- ----------
Cor13 ---------- ---------- ---------- ---------- ----------
Cor3  ---------- ---------- ---------- ---------- ----------
Cor1  ---------- ---------- ---------- ---------- ----------
Cor10 ---------- ---------- ---------- ---------- ----------
Cor14 ---------- ---------- ---------- ---------- ----------
Cor12 ---------- ---------- ---------- ---------- ----------
Cor15 ---------- t--------- ---------- ---------- ----------
Cor4  ---------- ---------- ---------- ---------- ----------
Cons. TTATATGCAA CAGGAGAAAT AATAGGAGAA ATAAGACAAG CATATTGTAA
```

Figure 4.1 (cont).

```
    251                                                        300
Cor5  ---------- ---------- ---------- ---------- ----------
Cor6  ---------- ---------- ---------- ---------- ----------
Cor7  ---------- ---------- ---------- ---------- ----------
Cor8  ---------- ---------- ---------- ---------- ----------
Cor2  ---------- ---------- ---------- ---------- ----------
Cor9  ---------- ---------- ---------- ---------- ----------
Cor11 ---------- ---------- ---------- ---------- ----------
Cor13 ---------- ---------- ---------- ---------- ----------
Cor3  -----a---- ---------- ---------- ---------- ----------
Cor1  ---------- ---------- ---------- ---------- ----------
Cor10 ---------- ---------- ---------- ---------- ----------
Cor14 ---------- ---------- ---------- ---------- ----g-----
Cor12 ---------- ---------- ---------- ---------- ----------
Cor15 ---------- ---------- ---------- ---------- ----------
Cor4  ---------- ---------- ---------- ---------- ----------
Cons. CCTTAGTAGA ACACAATGGA AGGACACTTT AAAACAGATA GTTATAAAAT

     301                                                        350
Cor5  ---------- ---------- ---------- ---------- ----------
Cor6  ---------- ---------- ---------- ---------- ----------
Cor7  ---------- ---------- ---------- ---------- ----------
Cor8  ---------- ---------- ---------- ---------- ----------
Cor2  ---------- ---------- ---------- ---------- ----------
Cor9  ---------- ---------- ---------- ---------- ----------
Cor11 ---------- ---------- ---------- ---------- ----------
Cor13 ---------- ---------- ---------- ---------- ----------
Cor3  ---------- ---------- ---------- ---------- ----------
Cor1  ---------- ---------- ---------- ---------- ----------
Cor10 ---------- ---------- ---------- ---------- ----------
Cor14 ---------- ---------- ---------- ---------- ----------
Cor12 ---------- ---------- ---------- ---------- ----------
Cor15 ---------- ---------- ---------- ---------- ----------
Cor4  ---------- ---------- ---------- ---------- ----------
Cons. TAAGAGAACA ATTTAAGAAT AAAACAATAG TCTTTAATCA ATCCTCAGGA

     351                  373
Cor5  ---------- ---------- ---
Cor6  ---------- ---------- ---
Cor7  ---------- ---------- ---
Cor8  ---------- ---------- ---
Cor2  ---------- ---------- ---
Cor9  ---------- ---------- ---
Cor11 ---------- ---------- ---
Cor13 ---------- ---------- ---
Cor3  ---------- ---------- ---
Cor1  ---------- ---------- ---
Cor10 ---------- ---------- ---
Cor14 ---------- ---------- ---
Cor12 ---------- ---------- ---
Cor15 ---------- ---------- ---
Cor4  ---------- ---------- ---
Cons. GGGGACCCAG AAAGCTTGTG TAC
```

Figure 4.1 (cont).

There were 10 nucleotide changes distributed between 7 individual variants, corresponding to a mutation frequency of 1.7 X 10^{-3}/nucleotide. However, no clones showed any deletions or insertions when compared with each other or with the HIV-1 prototype sequence of strain HXB2.

Translation of the DNA sequences to their respective protein sequences showed that 4/10 of the nucleotide changes were silent (Figure 4.2). The expected frequency of silent changes from random occurrence is about 20%-30% (Jukes et al., 1979), although this figure depends on the base composition of the sequence. The high frequency of silent mutations (40%), is consistent with what might be expected in the absence of selection pressures for amino acid changes. In other studies investigating the sequence variation of *env* at later stages of HIV-1 infection (Pang et al., 1991; Simmonds et al., 1990; Balfe et al., 1990; Wolfs et al., 1991), the low frequency of silent mutations (1%-8%)suggested that selection pressures for amino-acid changes were playing a role in the emergence of particular HIV-1 variants. In the present chapter, and in the study of Pang and colleagues (1992) (Pang et al., 1992), the increased incidence of silent mutations indicates that during early times of HIV-1 infection, selection, at least in the peripheral circulation, is not yet playing a major role in emergence of variants within either the V3 region (present study) (Ait-Khaled and Emery, 1993),or within the V4 region (Pang et al., 1992). The multiple alignment of the protein sequences is shown in Figure 4.2.

```
       1                                                  50
cln.1  ---------- ---------- ---------- ---------- ---------K
cln.3  ---------- ---------- ---------- ---------- ----------
cln.4  ---------- ---------- ------D--- ---------- ----------
cln.14 ---------- ---------- ---------- ---------- ----------
cln.15 --------I- ---------- ---------- ---------- ----------
Cons.  LGIRPVVSTQ LLLNGSLSEG EVVIRSSNFT DNAKIIIVQL NKSVEINCTR

       51                                                 100
cln.1  ---------- ---------- ---------- ---------- ----------
cln.3  ---------- ---------- ---------- ----N----- ----------
cln.4  ---------- ---------- ---------- ---------- ----------
cln.14 ---------- ---------- ---------- ---------- -------R--
cln.15 ---------- ---------I ---------- ---------- ----------
Cons.  PNNNTRKSIH IGPGKTLYAT GEIIGEIRQA YCNLSRTQWK DTLKQIVIKL

      101                    123
cln.1  ---------- ---------- ---
cln.3  ---------- ---------- ---
cln.4  ---------- ---------- ---
cln.14 ---------- ---------- ---
cln.15 ---------- ---------- ---
Cons.  REQFKNKTIV FNQSSGGDPE SLC
```

Figure 4.2: Multiple alignment of the V3 amino acid sequences derived from the DNA alignment shown in figure 4.2. Five variants have unique sequences whereas the remaining 10 were identical to the consensus sequence. The V3 loop is underlined.

The 6 coding base substitution, resulted in only 5 amino acid changes because one of the aa change was the result of 2 bases substitutions which is a rare event. Five variants possessed amino acid alterations with two species exhibiting variants within the V3 loop (clone #1: aa 298 R6K; clone #15: aa 318 T6I). Comparison of the V3 consensus with the prototype HIV-1 *env* V3 sequence of strain HXB2 reveals many variations especially within the V3 loop 3' of the GPG motif, however the overall charge associated with the loop is maintained at +3. Moreover, the substitutions for acidic aa residues at the specific positions involved in the SI phenotype, namely, positions 306, 320, 324 substituting S to R, D to Q and D to N respectively are not observed in any of the 15 seroconversion variants. This is in accordance with the majority of similar studies showing that the transmitted virus is of an NSI phenotype. In addition, the GPGKTL hexapeptide is not present in any of the HIV-1 *env* V3 sequences currently deposited in the EMBL database (release 33).

In order to ascertain whether these alterations are likely to affect the overall structure of the V3 loop, the predicted secondary structures of the V3 loop from the variants identified in this study and the prototype HXB2 V3 sequence have been compared. The results indicated that the overall structure of the loop is maintained including the ß-turn at the tip of the loop (data not shown).

4.2.2 Sequence analysis HIV-1 V3 region natural variants present in the lymph node and peripheral blood of Patient 1

The HIV-1 V3 region from proviral DNA present in the lymph node and PBMCs of patient 1 has been amplified by PCR, cloned into the Sma I restriction site of pUC18 and DNA sequences from the resultant clones were determined. A total of 12 blood variants and 11 lymph node variants were sequenced. Representative autoradiographs of sequencing polyacrylamide gels are shown in Figures 4.3 to 4.5.

The 23 DNA sequences were aligned using multiple pairwise alignment as shown on Figure 4.6. The consensus sequence was derived from the multiple alignment. For comparison, the consensus sequence of the V3 region of patient 0 (figure 4.1) is also shown in figure 4.7. Interestingly, differences between variants are not homogeneously distributed along the V3 region analysed. Only 24 distinct point mutations were observed between bases 1 to 150, coresponding to aa 250 to 295 (5' of the V3 loop). In contrast, 33 distinct point mutations are noted between bases 250 and 350 as well as a 9 base-pair insertion in 7 variants (Vb7, Vln6, Vb3, Vln9, Vb11, Vln4, and Vln5), this region is located 3' of the V3 loop (aa 296 to 330). The region encompassing the V3 loop, between bases 150 and 250, showed the lowest heterogeneity, with only 10 distinct point mutations amongst the 23 variants.

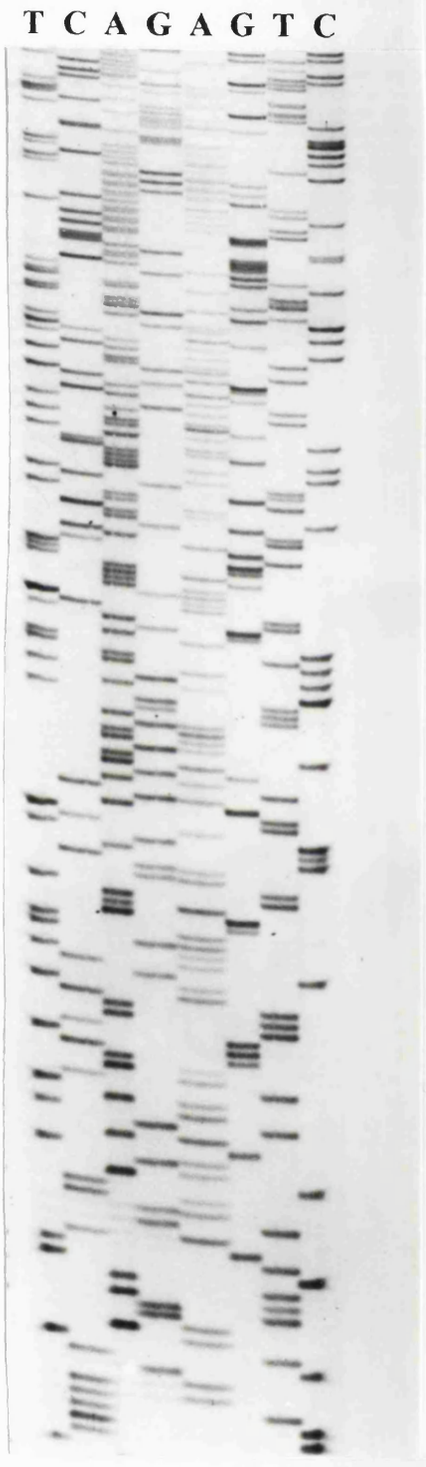


Figure 4.3: Autoradiograph of a proviral V3 region variant (Vb1) from patient 1.

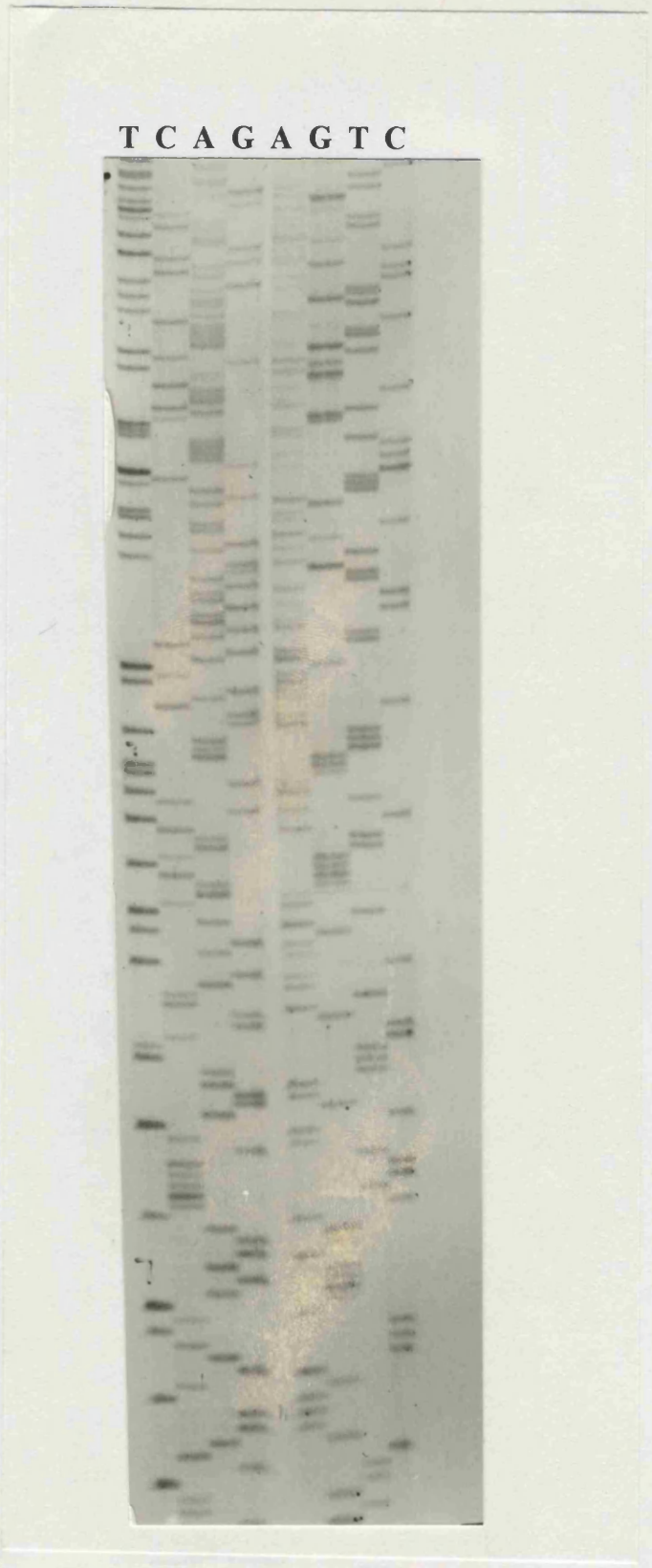


Figure 4.4: Autoradiograph of a proviral V3 region variant (Vln7) from patient 1.

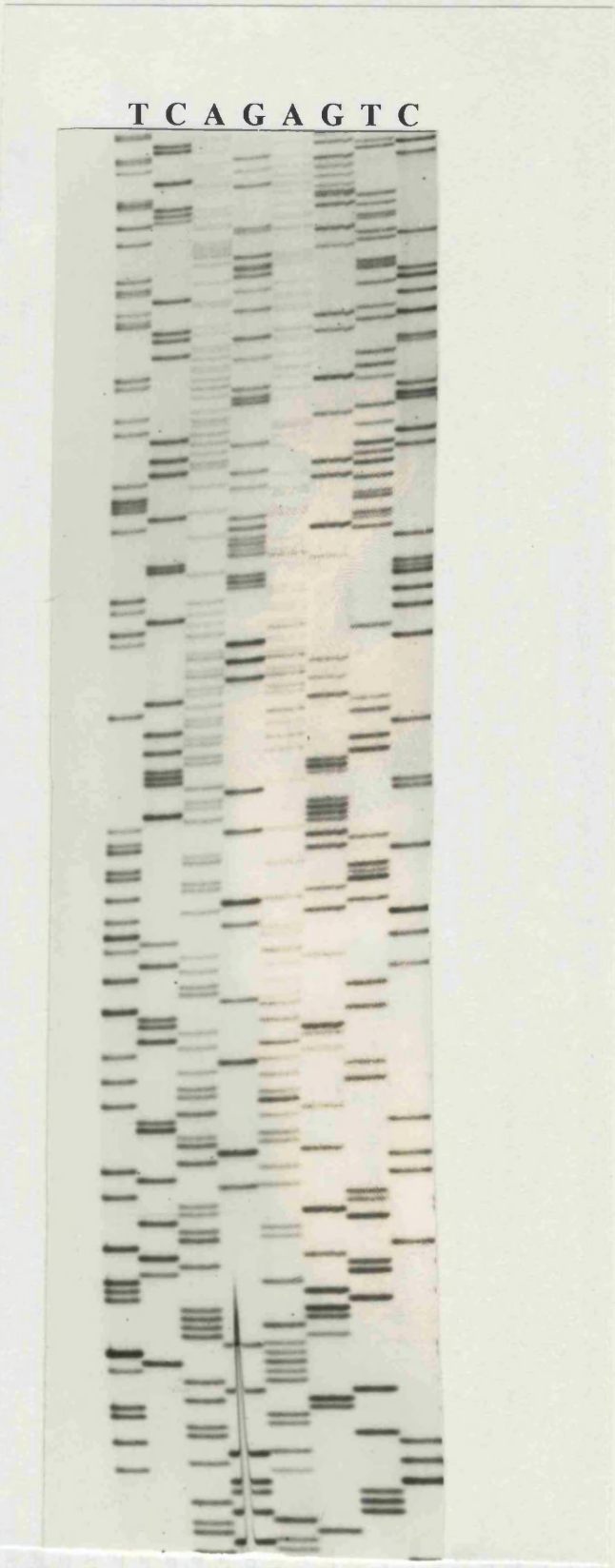


Figure 4.5: Autoradiograph of a proviral V3 region variant (Vln3) from patient 1.

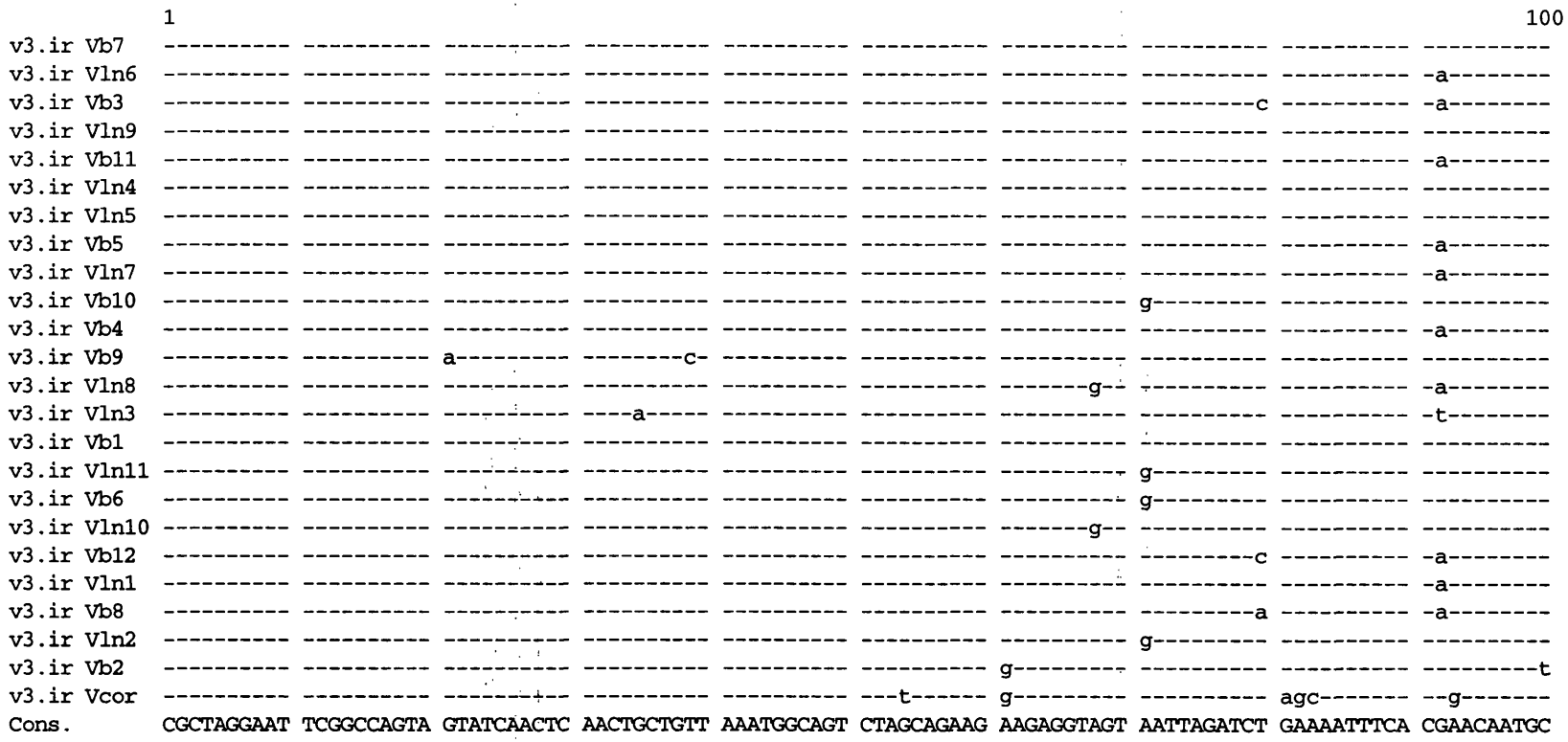


FIGURE 4.6: Multiple sequence alignment of V3 region proviral clones isolated from blood (v3.ir Vb1 to 11) and from a lymph node of patient 1 (v3.ir Vln1 to 11). The consensus sequence derived from the alignment (cons.) is alao shown. The consensus sequence of patient 0 blood isolates (v3.ir Vcor) is shown for comparison. Deletions are indicated as points, and identity to the consensus sequence are shown as dashes.

```
          101                                                 150                                                 200
v3.ir Vb7   -----a---- --------c- ---------- -----t---- ---------- ---------- ---------- ---------- ----------
v3.ir Vln6  -----a---- ---------- ---------- -----t---- ---------- ---------- ---------- ---------- ----------
v3.ir Vb3   ---------- ---------- ---------- ---------- ---------- ---------- ---c------ ---------- ----------
v3.ir Vln9  ---------- ---------- ---------- -----g---- ---g------ ---------- ---c------ ---------- ----------
V3.ir Vb11  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---c------
v3.ir Vln4  -----t---- ---------- ---------- -----t---- ---------- ---------- ---------- ---------- --c-a-----
v3.ir Vln5  ---------- ---------- ---------- ---------- ---------- ----t----- ---------- ---------- ----------
v3.ir Vb5   -----t---- ---------- ---------- ---------- ---a------ ---------- ---------- ---------- ----------
v3.ir Vln7  -----t---- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
v3.ir Vb10  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
v3.ir Vb4   -----t--c- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
v3.ir Vb9   -----t---- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
v3.ir Vln8  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
v3.ir Vln3  --g------- --------c- ---------- -----t---- ---------- ---------- ---------- ---------- ----------
v3.ir Vb1   ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
v3.ir Vln11 ---------- ---------- ---------- ---------- --------c- ---------- ---------- ----a----- ----------
v3.ir Vb6   ---------- ---------- ---------- ---------- ---------- ---------- ---------- -a-------- --c-a-----
v3.ir Vln10 ---------- --------t- ---------- ---------- ---c------ ---------- ---------- -a-------- --c-a-----
v3.ir Vb12  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
v3.ir Vln1  ---------- --------c- ---------- -----t---- ---------- ---------- ---------- ---------- ----------
v3.ir Vb8   -----t---- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
v3.ir Vln2  -----a---- ---a------ -a-------- ---------- ---------- ---------- ---------- ------g--- ----------
v3.ir Vb2   ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ------a---
v3.ir Vcor  -----t---- ---------- ------a--- ---------- ---------- ---------- ---------- -a-------- ----a----- a--c-aaa--
Cons.       TAAAACCATA ATAGTACAGC TGAATACATC TGTAGAAATT AATTGTACAA GACCCAACAA CAATACAAGA AGAAGTATAC ATATGGGACC GGGGAGTGCA
```

Figure 4.6 (cont).

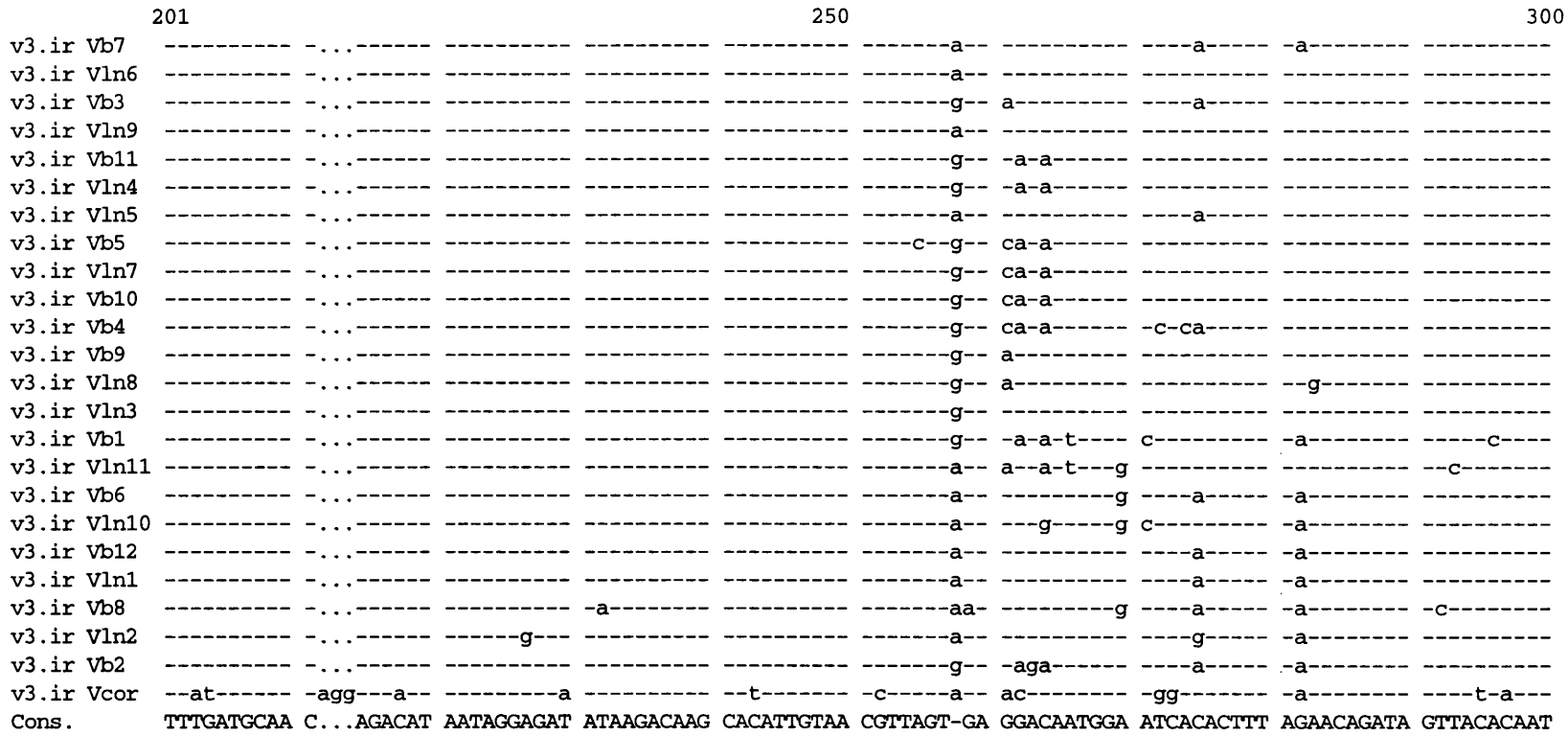


Figure 4.6 (cont).

```
          301                                                                                                  350
v3.ir Vb7   ---------- a------g-- ---------- ---------- --c------g ---------- ---------- ---------- -----
v3.ir Vln6  ---------- a--------- ---------- ---g------ --c------g ---------- ---------- ---------- -----
v3.ir Vb3   ---------- a--------- ---------- ---------- --c------g ---------- ---------- ---------- -----
v3.ir Vln9  ---------- a--------- ---------c ---------- --c------g ---------- ---------- ---------- -----
v3.ir Vb11  ----g----- a--------- gat------- ---------- ---------g ---------- ---------- ---------- -----
v3.ir Vln4  ---------- a--------- gat------- ---------- ---------- ---------- ---------- ---------- -----
v3.ir Vln5  ---------- g--------- gat------- ---------- ---------- ---------- ---------- ---------- -----
v3.ir Vb5   ---------- g--------- .........- ---------- -a-------- ---------- ---------- ---------- -----
v3.ir Vln7  ---------- g--------- .........- ----g----- -a-------- ---------- ---------- ---------- -----
v3.ir Vb10  ---------- g--------- .........- ---------- -a-------- ---------- ---------- ---------- -----
v3.ir Vb4   ---------- g--------- .........- ---------- -a-------- ---------- ---------- ---------- -----
v3.ir Vb9   ---------- a--------- .........- ---g------ ---------- ---------- ---------- ---------- -----
v3.ir Vln8  ---------- a--------- .........- ---g------ ---------- ---------- ---------- ---------- -----
v3.ir Vln3  ---------- a--------- .........- ---------- -------g-- ---------- ---------- ---------- -----
v3.ir Vb1   --g------- a--------- .........- ---g------ ---------- ---------- ---------- ---------- -----
v3.ir Vln11 ---------- a--------- .........- ---------- ---------- ---------- ---------- ---------- -----
v3.ir Vb6   --g------g g--------- .........- ---g------ ---------- ---------- ---------- ---------- -----
v3.ir Vln10 ---------- g--------- .........- ---------- ---------- ---------- ---------- ---------- -----
v3.ir Vb12  ---------- g--------- .........- ---------- -------g-- ---------- -------ttt ---------- -----
v3.ir Vln1  ---------- g--------- .........- ---------- ---------- ---------- ---------- ---------- -----
v3.ir Vb8   ---------- g--------- .........- ---------- --c------- ---------- ---------- ---------- -----
v3.ir Vln2  ---------- g--------- .........- ---------- ---------- ---------- ---------- ---------- -----
v3.ir Vb2   ----g----- g----a-... .........- ---------- ---------- ---------- ---------- ---------- -----
v3.ir Vcor  -----g---- a---aa-... .........- ---------- ---------- ---------- ---------- ---------- -----
Cons.       TAAGAAAACA -TTTGGGAAT ACATTGGGGA ATAAAACAAT AGTCTTTAAT CAATCCTCAG GAGGGGACCC AGAAAGCTTG TGTAC
```

Figure 4.6 (cont).

Translation of the DNA sequences to their respective protein sequences (Figure 4.7) showed only 5 amino acid mutations in the V3 loop (aa296 to aa330) . Between aa250 and aa295 (5' of the V3 loop), 9 amino acid mutations were noted. The highest number of coding changes were between aa331 and aa375 (3' of the V3 loop), with 29 distinct amino acid differences amongst the 23 variants, as well as a 3 amino acid insertion. This high frequency of coding changes in the V3 region located 3' of the V3 loop (88%) is much higher than what would be expected based upon random occurrence, and suggests that selective pressures for amino-acid changes were playing a role in the emegence of particular variants. Inversely, the lower frequencies of coding changes in the region 5' of the V3 loop (38%), and in the V3 loop itself (50%), indicate that these regions of the third variable region of HIV-1 envelope must be under stronger functional constraint selection than the region located 3' of the V3 loop.

The multiple sequence alignment shown in Figure 4.6 was used as the basis to calculate the genetic distance between variants in each compartment. There was no significant difference in the heterogeneity of the lymph node and simultaneous blood quasispecies; variants from both compartments had a median divergence of 3%.

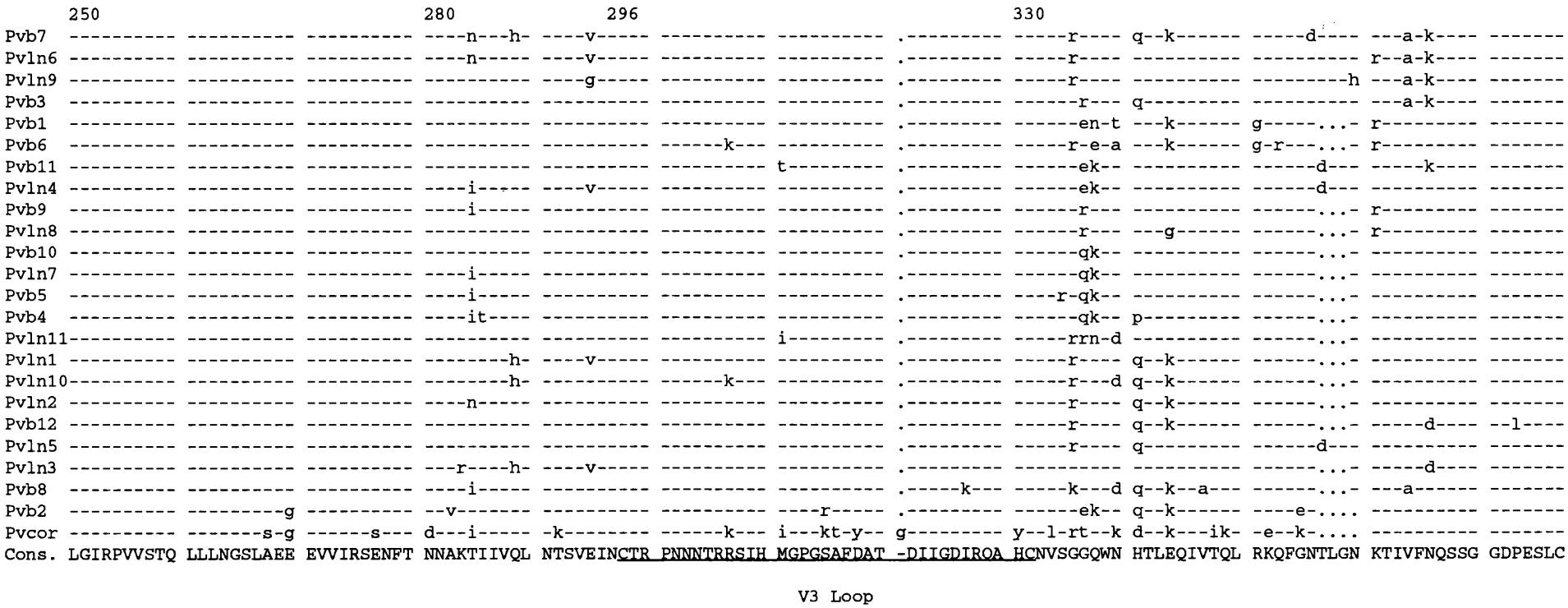


Figure 4.7: Alignment of V3 region amino acid sequences (aa 249 to 375 based on Hxb2 clone) of patient 2 blood variants (Pvb1 to Pvb11) and lymph node variants (Pvln1 to Pvln11) with the consensus sequence (cons.). The consensus sequence of patient 0 blood isolates (Pvcor) at seroconversion is shown for comparison. Gaps are indicated as points; similarity to the consensus sequence is shown as dashes. The V3 loop (aa296 to 330) is underlined. Patient 1 V3 loop is characterized by aa 319 deletion, and a GPGS central motif or PND.

4.2.3 Phylogenetic analysis of V3 region nucleotide sequences from patient 1

To further investigate the relationship between the lymph node and blood quasispecies, the sequences were analysed by phylogenetic methods (section 2.10). The consensus maximum parsimony tree (Figure 4.8) of the DNA sequences showed a mixed population with blood and lymph node variants in common branches. Branches with short inter-node branch length (low bootstrap values) are characteristic of very closely related variants (VLN3, VLN8, VB9, VB1, VB1, VLN2, VB8, VB12, VLN1, VLN11, VLN10, and VB6). The tree also reveals two other distinct clades, one consisting of VLN4, VB11, VLN5, VB7, VLN6, VLN9, and VB3, with a bootstrap value of 97%; the other clade comprises VB10, VLN7, VB5, and VB4, and has a bootstrap value of 68%.

4.2.4 Phylogenetic analysis of the V3 region amino acid sequences from patient 1

Similarly, a maximum parsimony tree was constructed for the amino-acid sequences of the blood and lymph node variants of patient 1 (Figure 4.9). Interestingly, the two distinct clades described for the proviral DNA tree, are exactly reproduced in the phylogenetic tree of the amino acid sequences, indicative of the fact that most of the point mutations of the proviral variants resulted in amino acid changes. These 2 clades consist of the corresponding amino acid sequences, PVLN4, PVB11, PVLN5, PVB7, PVLN6, PVLN9, and PVB3, with a bootstrap value of 77%, and PVB10, PVLN7, PVB5, and PVB4, with a bootstrap value of 72.7%. The other protein variants are closely related to each other and are associated with low bootstrap values.

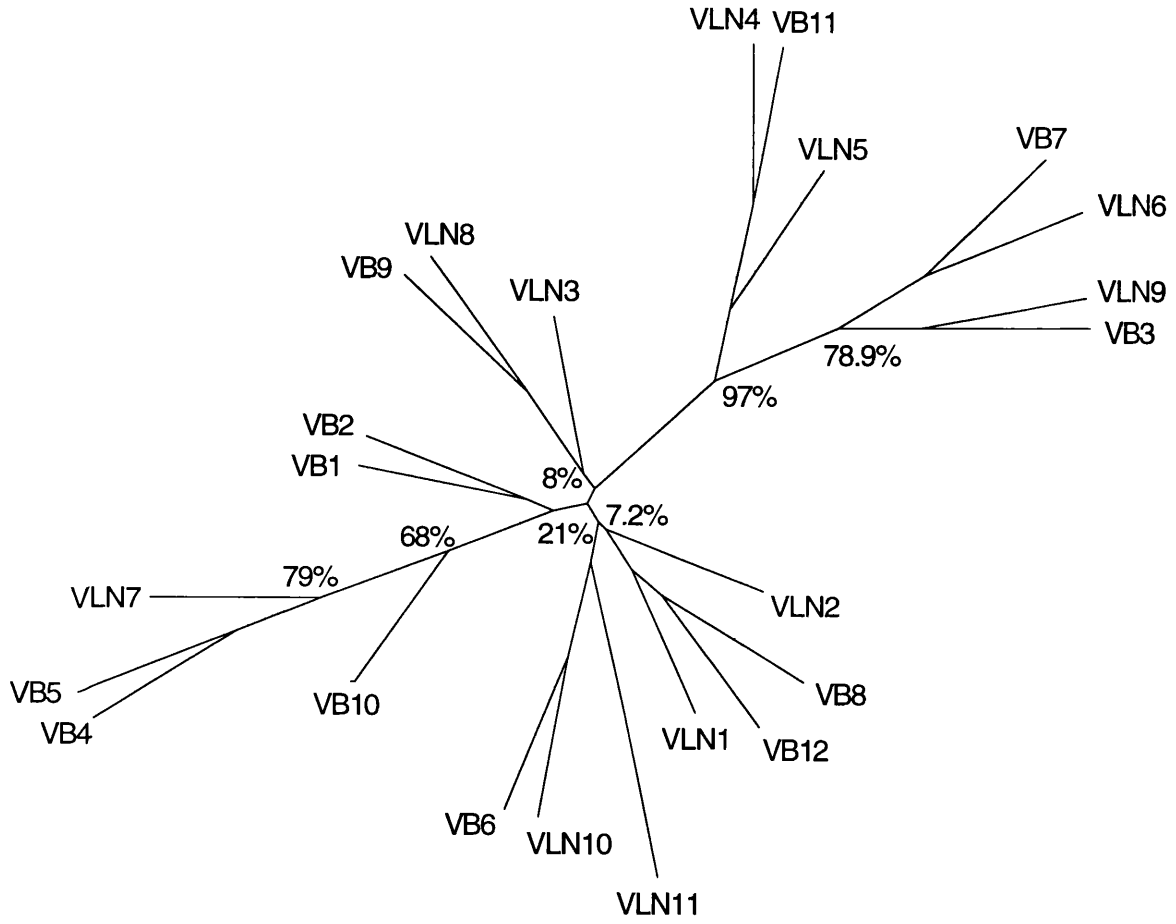


Figure 4.8: Maximum parsimony tree depicting the phylogenetic relationship of the 12 blood proviral sequences of the V3 region (VB1 to VB12) and the 11 lymph node variants (VLN1 to VLN11). All branches are drawn to scale, and the relevant bootstrap values indicated at the nodes of the branches.

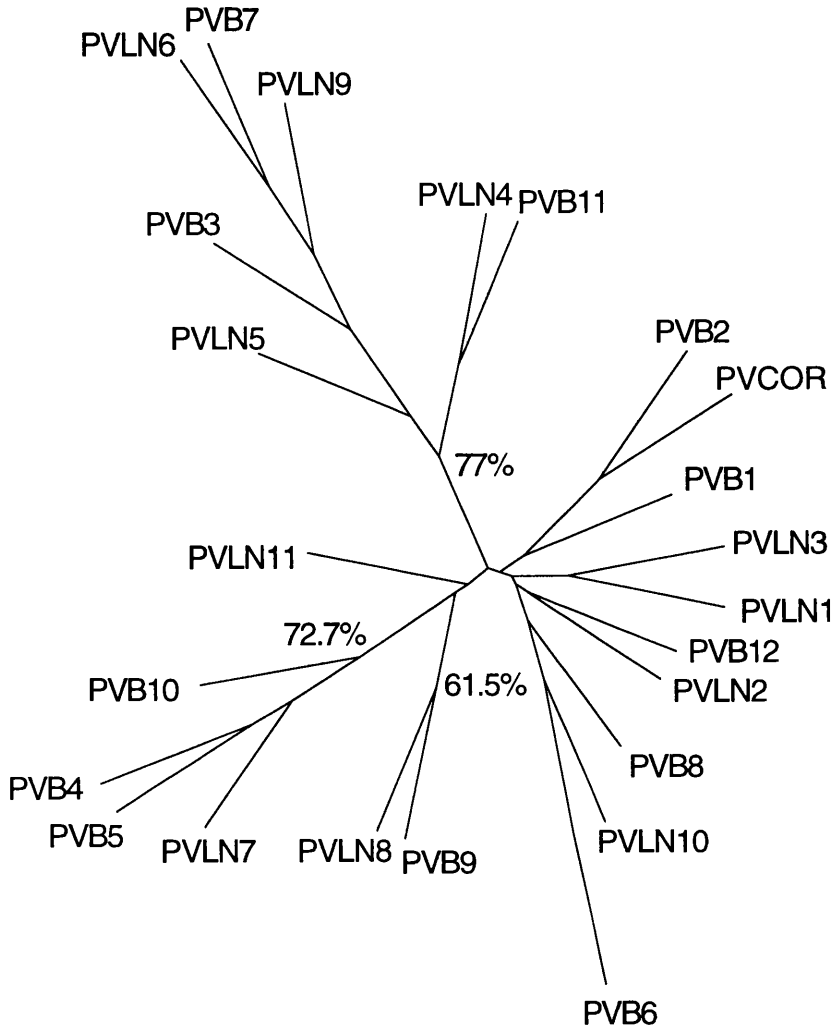


Figure 4.9: Maximum parsimony tree depicting the phylogenetic relationship of the 12 blood amino acid sequences of the V3 region (PVB1 to PVB12) and the 11 lymph node variants (PVLN1 to PVLN11). All branches are drawn to scale, and the relevant bootstrap values indicated at the nodes of the branches.

Bearing in mind that the V3 region is a major determinant of tropism, the sequence similarities between the blood and lymph node V3 variants is in agreement with the fact that the major target cells for HIV-1 are the CD4 T lymphocytes in both blood and lymph node. The phylogenetic tree also showed that there was not any compartmentalised evolution of HIV-1 variants in the two body sites investigated, suggesting that variants from these sites probably became mixed by transit of lymphocytes through lymphoid tissue, and seeding of HIV-1 infected cells from the histologically impaired lymph node to the circulation.

4.3 DISCUSSION

The distribution of HIV-1 V3 variants present in the peripheral blood has been determined in an individual with documented seroconversion to HIV-1 and in an AIDS patient. A striking feature of the data obtained is the homogeneity throughout the V3 region at both the proviral DNA and protein levels of the quasispecies from the seroconverter, contrasting with the extensive diversity observed in the blood and lymph node variants obtained from the AIDS patient. In the case of patient 0, there were only 10 mutations present within the 15 clones sequenced with one major quasispecies predominating at the nucleic acid level (53%) and at the protein level (66%). It is interesting to note that the conserved central motif within the V3 loop sequence present in this seroconverter was relatively divergent from the majority of V3 sequences available in the databases, and had a GPGKTL sequence at the crown of the loop rather than the more common GPGRAF. At present, no X-ray diffraction data are available for *env*, although nuclear magnetic resonance (NMR) solution structures of synthetic V3 loop peptides confirm the presence of a ß-turn at the crown of the loop (Chandrasekhar et al., 1991; Zvi et al., 1995b; Zvi et al., 1995a). The secondary structure predictions used in this study are compatible with the solution structures and showed that the overall consensus structure of the V3 loop is likely to be maintained in the quasispecies.

Genetic homogeneity within the V3 loop at the time of seroconversion has been described elsewhere (Zhang et al., 1993; McNearney et al., 1992; Wolfs et al., 1992; Zhu et al., 1993). These studies indicate that a single variant is present at seroconversion which is closely related to the prototype HIV strains. Whilst the data presented here show that the V3 region and V3 loop are relatively homogeneous, other minor quasispecies were apparent. Whether these quasispecies reflect the

time at which the sample was obtained ie HIV antigen positive and HIV antibody equivocal, is difficult to ascertain. However, the presence of minor strains within the V3 loop at such early stages of infection indicates the rapidity with which HIV quasispecies is generated in the absence of humoral immune selection. Similar data have been described by Pang et al (1992) (Pang et al., 1992) in two seroconverters when analysing V4 quasispecies. Interestingly in the study of Pang and co-workers, the investigators analysed two longitudinal samples from a seroconverter, and observed variation in V4 only in the second sample (10 species out of 30 clones). In the first blood sample obtained 5 days before the second sample, there was no variation in the 30 HIV-1 V4 region variants sequenced (30 identical clones) (Pang et al., 1992). This finding supports *de novo* variation, although the possibility that those variants were present in the first sample but at levels too low to be detected cannot be discounted.

Since the amino acid variation that contribute to the syncytium inducing (SI) phenotype have been described (de Jong et al., 1992b), the V3 loop sequence has been analysed for the presence of such mutations. None of the variants contained alterations at any of the amino acids shown to contribute to the SI phenotype (#317, #320 and #324), and the overall charge of the loop remains the same as the non syncytium inducing (NSI) phenotype (net charge +3). It is likely therefore that the HIV strains in this individual are of the NSI phenotype at this stage of infection.

The other factor to be considered in these analyses is the relative proportions of quasispecies present and how they relate to the original inoculum. Two possibilities exist:

1- The distribution of quasispecies within V3 is closely related to those strains

present in the infecting inoculum without any selection effect.

2- Functional constraints for the infection of CD4+ T-cells and cells of the monocyte/macrophage lineage result in the outgrowth of HIV-1 strains with V3 loop sequences that can efficiently establish infection within these cells.

The first option appears unlikely on the basis of data demonstrating a rapid evolution of HIV-1 envelope sequences throughout disease (see Section 1.3) and may only be valid if the inoculum was derived from a seroconverter. The second option appears more tenable and would concur with population biology views on host-parasite interactions (McLean and Nowak, 1992). Subsequent to initial infection, a plethora of strains evolve to fill their respective anatomical sites such as brain, spleen, and attempt to evade both B and T-cell immune surveillance. It should be noted that although sequence variation in *env* is important for cell tropism, other factors such as the presence, absence, or modification of transcription factor binding sites within the LTR could have a profound effect on the replicative competence of viruses in particular cell and organ types (see Section 1.3.4 and chapter 6).

In conclusion, the data presented suggest that at early times after infection with HIV, sequence evolution has already started and these strains are present as integrated proviral DNA in the host. However, a relatively homogeneous collection of V3 variants are present at this early stage of infection. Moreover, these variants are of the NSI phenotype as predicted from the V3 loop sequences.

In contrast to the V3 quasispecies from patient 0 (seroconverter), that from patient 1 (Center for Disease Control (CDC) IV stage), was more heterogeneous. Alignment of the derived amino acid sequence of the 23 V3 region (aa249 to aa375) variants

(11 lymph node clones and 12 blood clones), showed no identical sequence, all sequences differed from each other by at least one amino acid substitution. There was no significant difference in the heterogeneity of the lymph node and simultaneous blood quasispecies; variants from both compartment had a median divergence of 3%. Despite the relatively low CD4 count of patient 1 (220 x 10^6/l) and CDC IV disease, all the V3 variants analysed in both lymph node and blood were predicted to be of the NSI phenotype. Indeed, all but one variant, B2, had an overall charge of +3 and would be expected to be of the NSI phenotype. The fixed substitution of acidic amino acids at positions 306, 317, 320 and 324 associated with an SI phenotype were not present in any of the V3 variants (Figure 4.7). A single V3 amino acid sequence (B2) had an overall charge of +4, and would be predicted to be an intermediate SI variant. The predominance of V3 variants of predited NSI phenotype in this AIDS patient is in accordance with the fact that no rapid CD4 count decline below 200 was observed in this patient 4 months following biopsy and blood sampling as might be expected if SI variants were present (Richman and Bozette 1994). The V3 loop of patient 1 was characterised by aa G319 deletion and a GPGS central motif, which like that of patient 0 an uncommon central motif (LaRosa et al., 1990).

Interestingly, the heterogeneity of the V3 region was not evenly distributed and based on the frequencies of coding to non-coding mutations, three distinct regions of the sequence could be identified. The V3 region located 3' of the V3 loop showed a high frequency of coding changes(88%) suggesting that selective pressures for amino-acid changes were playing a role in the emegence of these variants. Inversely, the lower frequencies of coding changes in the region 5' of the V3 loop (38%), and in the V3 loop itself (50%), indicate that these regions of the third

variable region of HIV-1 envelope must be under stronger functional constraint selection than the region located 3' of the V3 loop. Overall, the high rate of non-synonymous nucleotide substitution in the V3 region indicates that strong positive selection for amino acid change is operating in the generation of antigenic diversity.

The phylogenetic tree also shows that there was not any compartementalised evolution of HIV-1 variants in the two body sites investigated. These results do not invalidate the hypothesis involving the lymph node as a reservoir of viral variants, where HIV-1 virions and HIV-1-infected cells are sequestred in the lymph node leading to compartmentalised evolution. Indeed, the dichotomy between lymph node and blood regarding viral load levels (Pantaleo et al., 1993a; Embretson et al., 1993), as well as viral phenotypes (Tamalet et al., 1994; Lafeuillade et al., 1995), requires an intact histological architecture of the lymph nodes. At later stages of HIV-1 infection, the histological structure of the lymph nodes are disrupted, including follicular lysis, and an absence of germinal centres and follicular dendritic cell network. The results presented in this chapter were obtained from proviral sequences derived from a lymph node biopsy which upon histological examination showed follicular lysis and disruption of the germinal centres. The role of the lymph nodes in compartmentalised evolution will be discussed further in chapter 5.

CHAPTER 5.

PHYLOGENETIC RELATIONSHIP BETWEEN HIV-1 LONG TERMINAL REPEAT NATURAL VARIANTS PRESENT IN THE LYMPH NODE AND PERIPHERAL BLOOD OF THREE HIV-1 INFECTED INDIVIDUALS

5.1 INTRODUCTION

The elegant work of Wain-Hobson, Leigh Brown and others has shown that HIV exists in infected individuals as a collection of strains or a quasispecies (see Chapter 4, Section 4.1) (Wain-Hobson, 1993). Although the temporal variation between proviral HIV strains in the blood and plasma virus strains has been extensively investigated (Simmonds et al., 1991; Wolfs et al., 1992), relatively little is known about the site(s) of strain evolution or the relationship of variants between various HIV infected organs. Analysis of the V3/V4 regions of the *env* gene in brain, spleen and peripheral blood mononuclear cells (PBMCs) has indicated that strain evolution in each infected organ proceeds in parallel possibly reflecting cell tropism of particular variants (Epstein et al., 1991; Pang et al., 1991). Recent data has continued and extended earlier work from Janossy and colleagues (Tenner-Racz et al., 1986) highlighting the importance of the lymphoid tissues, and particularly the lymph nodes in HIV pathogenesis (Pantaleo et al., 1993a; Embretson et al., 1993). Thus, proviral HIV DNA has been shown to be present in a high proportion of lymph node CD4+ lymphocytes and macrophages at all stages of HIV infection. Such a situation contrasts with proviral HIV load in the peripheral blood, where a low proportion of CD4+ lymphocytes are infected at early and intermediate stages of HIV disease, although a higher load in the periphery is observed at late stages of disease (Pantaleo et al., 1993b). This effect may be accounted for by the seeding of HIV strains from the lymph node to the periphery following the loss of an effective cellular and humoral immune system and/or extensive destruction of the lymph node architecture (see chapter 1, section 1.6).

Although previous studies on tissue tropism of HIV-1 have concentrated on variation of *env* sequences in different tissues, a major drawback for these analyses is that

env regions are under selective pressure from both the B and T-cell arms of the immune system and may exhibit convergent evolution (Holmes et al., 1992), thus complicating any phylogenetic analyses. In the light of this observation, I have analysed the long terminal repeat (LTR) region of the HIV genome, since it is not subject to an as extensive immune pressure as *env*. The LTR functions as the promoter-enhancer unit of HIV and comprises multiple functional elements controlling the differential and temporal expression of viral genes (Gaynor, 1992). These protein-binding sites are under the influence of cellular (N-6B, Sp1, AP1, TCF-1∀) and viral (*tat,nef*) factors (Gaynor, 1992). Several cellular transcription factors binding the HIV-LTR are induced by activation or differentiation of lymphocytes or monocytes, rendering HIV subject to many of the same regulatory signals involved in the control of cellular gene expression (see Chapter 1, Section 1.2.5). In addition to the enhancer-promoter region (-158 to -1), the upstream region of the LTR (-420 to -159) has been identified as a negative regulatory element (NRE) whilst the proximal region contains the tat responsive TAR sequence spanning positions +1 to +60. At present, HIV-1 LTR strain variation has mainly been analysed in the peripheral blood (see Section 1.3.4).

In this chapter, I report the phylogenetic analyses of HIV-1 LTR variants present simultaneously in the lymph node and peripheral blood mononuclear cells of three HIV-1 infected individuals, as well as in blood samples preceding biopsy for patient 1, and samples following biopsy for patients 1, 2, and 3 and relate the results to both proximal markers of disease, clinical stage of infection and lymph node histology.

5.2 HIV-1 LTR SEQUENCE VARIATION AND PHYLOGENETIC ANALYSIS

5.2.1 HIV-1 LTR cloning and sequencing

The HIV-1 LTR from proviral DNA present in the lymph node and peripheral blood mononuclear cells of three patients has been PCR amplified, cloned into the Sma 1 site of pUC18 and the DNA sequence of multiple clones determined (see chapter 3). In the case of patient 2, the DNA sequence of 20 clones from DNA extracts of lymph node (LN1 to LN 20), simultaneous blood (BL1 to BL20), and an 11 months follow-up blood sample (BF1 to BF20) were determined. In the case of patient 1, the DNA sequence of 11 and 10 clones respectively from both lymph node(LN1 to LN11) and a simultaneous peripheral blood sample (LB1 to LB10) were determined, as well as 9 clones from a blood sample obtained 14 months prior to the biopsy (LBB1 to LBB9), and a blood sample obtained 6 months after the biopsy (LBF1 to LBF9). In the case of patient 3, the DNA sequence of 10 clones from DNA extract of lymph node (LN1 to LN10), simultaneous blood sample (B1 to B10), and from a blood sample obtained 17 month after the biopsy (BF1 to BF10). Representative autoradiographs of the LTR sequencing are shown in figures 5.1, and 5.2.

Figure 5.1: Autoradiograph of LN20 LTR sequencing from patient 2.

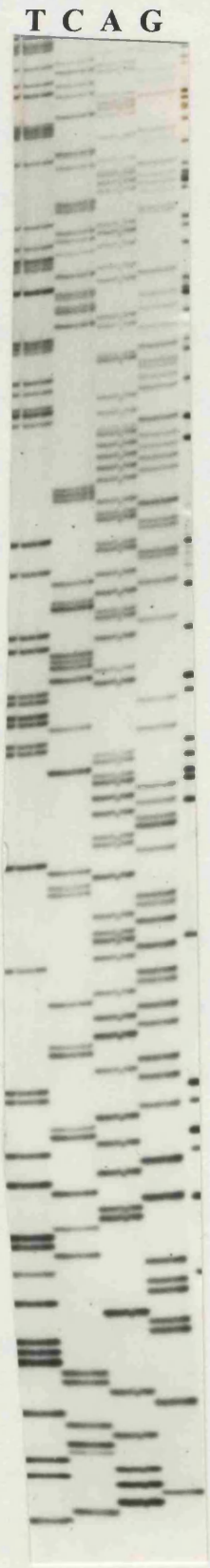


Figure 5.2: Autoradiograph of Lb9 LTR sequencing from patient 1.

5.2.2 HIV-1 LTR multiple alignments for patients 1, 2, and 3

The multiple sequence alignments were produced using multiple pairwise alignments for the 3 patients' LTR variants. Figures 5.3, 5.4, and 5.5 present these multiple alignments for patient 1, 2, and 3 respectively.

The distribution of mutations between variants was not even throughout the LTR, in particular, the region between positions +1 and +58 (relative to the start of transcription), containing the TAR sequence, showed almost no heterogeneity. The other transcription binding sites known to be crucial for HIV-1 replication, such as NF-6B and Sp1 sites, also showed little heterogeneity. Transcription factor binding sites thought to be of less importance, such as Site B, NF-AT, and TCF-1∀, dispayed the most heterogeneity.

In order to illustrate the pattern of mutations in the quasispecies, I will describe the multiple sequence alignment of the LTR variants isolated from patient 2 (Figure 5.4), in more detail. Within the characterised transcription factor binding site (-347 to -328 relative to the start of transcription based on HXB2 molecular clone), referred to as Site B (within the NRE), a major variant was identified. A T-cell protein recognizes the palindromic sequence within site B which also binds estrogen- or thyroid hormone-response elements with lower affinity, and Site B mutations results in increased expression from the HIV-1 LTR in T cell lines (Orchard et al., 1990). The variant identified in the quasispecies of patient 2 comprised 3 mutations; T 6 C at position -361 (inside the 5' protein binding site), C 6 T at position -355 and T 6 C at position -351 (inside the 9 bp spacer) (see figure 5.6). This variant was found in all LN clones except LN20, but in only 2 of the 20 BL clones (BL15 and17) and 3 of the BF clones (BF9, 12, and 13). It is interesting to note that these three

mutations were also detected in the quasispecies of patients 1 and 3, but not always linked.

Another feature of these variants, is a sequence duplication of 16 bp at position -134, 12 bases prior to the first NF-κB site and producing a second potential binding site for TCF-1α. All the BL variants possess this duplication, whereas, 19 of 20 BF variants, and 18 of the 20 LN also possess this duplication. The duplication was absent in variants LN1, LN10, and BF7. None of the 40 LTR variants of patient 1 possessed this duplication, and only 4 blood variants (B3, B4, B6, and B2) of patient 3 had a TCF-1α duplication. Another sequence duplication was present between the first two Sp1 sites in LN1 and BL10 variants of patient 2 resulting in the generation of a fourth Sp1 site in these variants. Neither patient 1 nor 3 harboured any variants with an Sp1 site duplication.

Finally, two mutations were detected in the *tat*-responsive element or TAR. Firstly, a C to T at position +24, within the bulge where *tat* binds, present in 19/20 BL variants, 12/20 LN variants, and 18/20 BF variants. This mutation was predominant in the quasispecies of patient 2 (consensus sequence of patient 1 variants in figure 5.4), and present in 9/30 of the natural variants of patient 3 (B3, B4, B6, B2, Ln6, Ln9, Ln7, Ln5, and Bf10; see figure 5.5). It is interesting to note that this mutation was present in an equal proportion of lymph node (4/10) and simultaneous blood variants (4/10), whereas only 1 of 10 follow up variants (Bf10) possessed the mutation. Secondly, a A to G mutation at position +47 (3' end of the stem) which is present in all the BL variants, 15/20 BF and 13/20 LN variants (see consensus sequence of patient 2, figure 5.4). This mutation was also present in 38/40 of patient 1 variants (see consensus sequence of patient 1 on figure 5.3), but only in 9/30 of

patient 3 variants (see consensus sequence of patient 3 on figure 5.5). These TAR mutations do not affect the secondary structure of the TAR RNA and have not yet been tested for their effect on *tat* transactivation (Roy et al., 1990; Harrich et al., 1995). However, uridine at position +23 (a T in the proviral alignment of figure 5.4), which is conserved in all variants, has been shown to be involved in a stable tertiary interaction with the GC pair directly above the bulge and therefore crucial for *tat* binding (Delling et al., 1992).The relatively high *in vivo* frequency of this mutation supports the fact that it does not affect *tat*-transactivation.

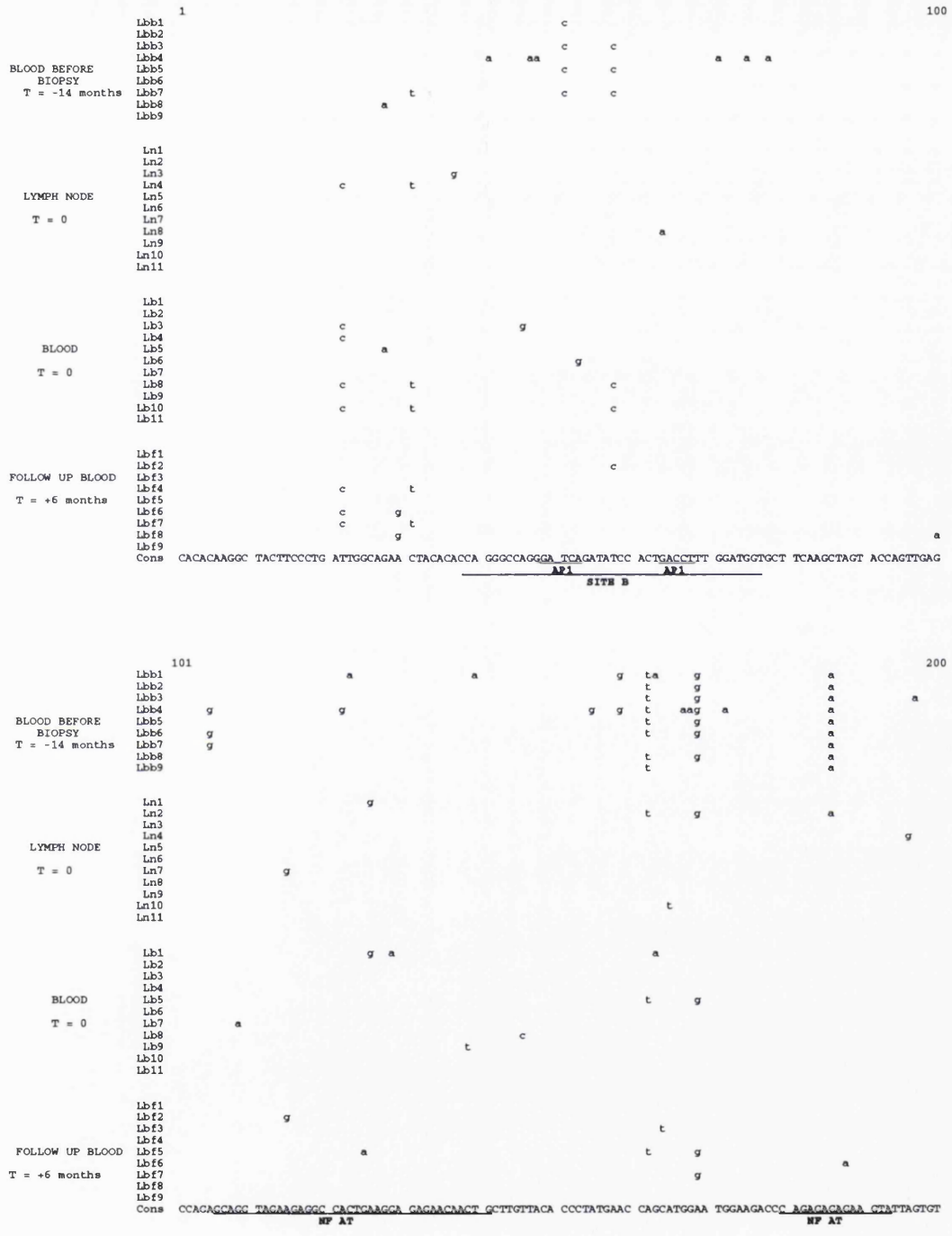


Figure 5.3: Multiple sequence alignment of the 40 LTR variants from patient 1. Ln, Lb, Lbb, and Lbf represent variants from lymph node,simultaneous blood, blood before biopsy, and following biopsy respectively. Points represent deletions. Transcription factor binding sites are underlined and annotated.

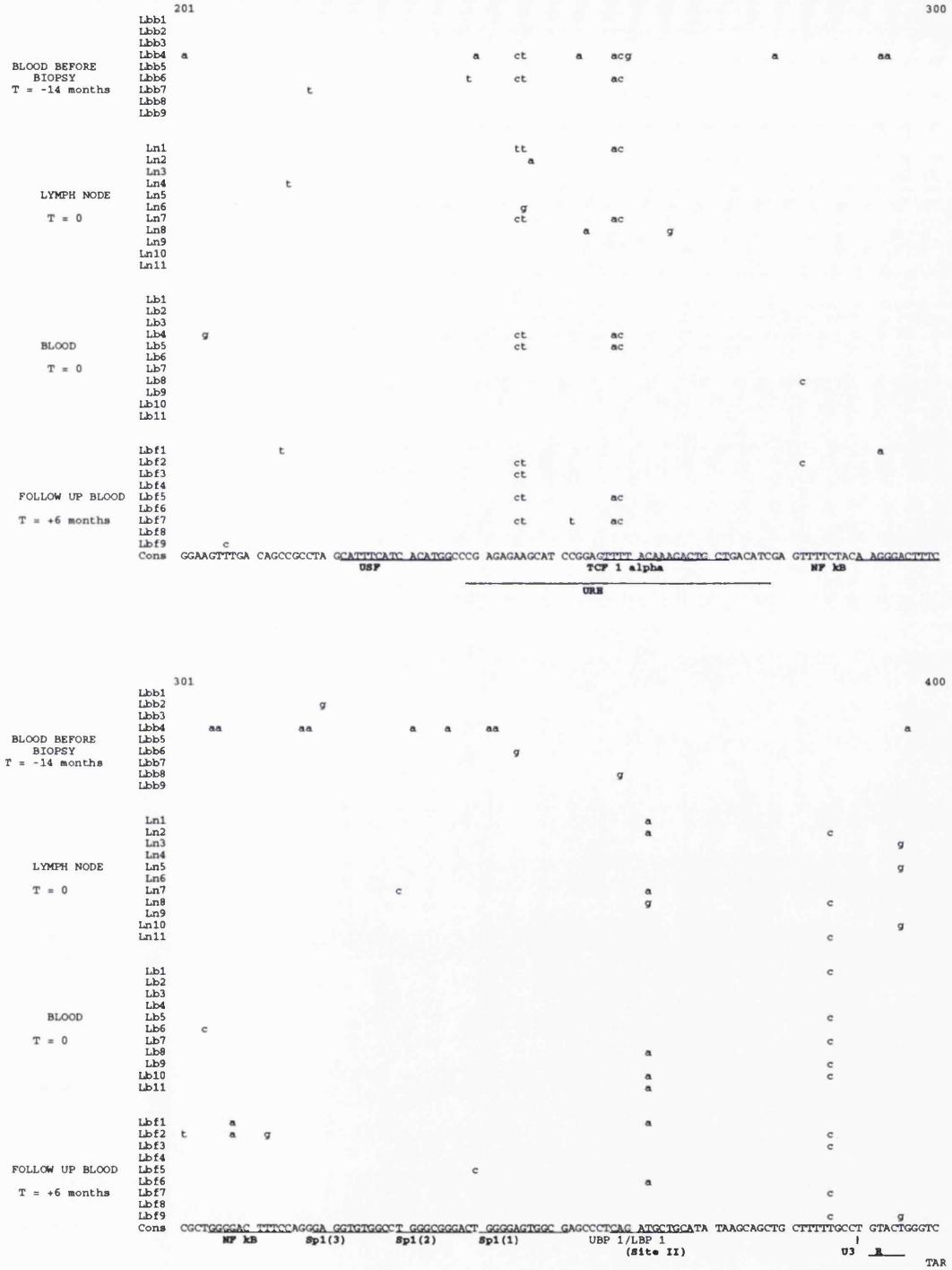


Figure 5.3 (cont): Multiple sequence alignment of the 40 LTR variants from patient 1. Ln, Lb, Lbb, and Lbf represent variants from lymph node,simultaneous blood, blood before biopsy, and following biopsy respectively. Points represent deletions. Transcription factor binding sites are underlined and annotated.

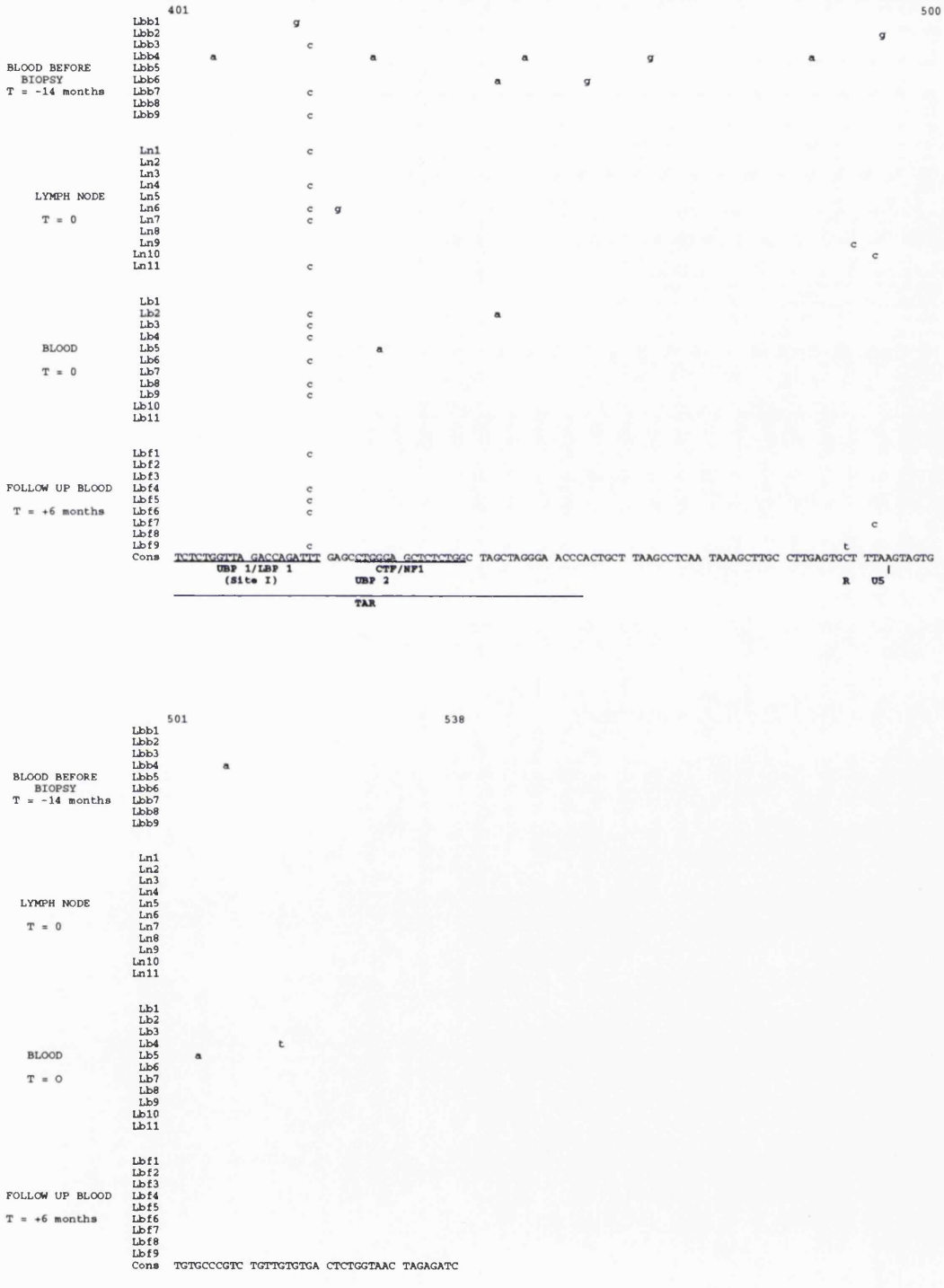


Figure 5.3 (cont): Multiple sequence alignment of the 40 LTR variants from patient 1. Ln, Lb, Lbb, and Lbf represent variants from lymph node,simultaneous blood, blood before lymph node biopsy, and following biopsy respectively. Points represent deletions. Transcription factor binding sites are underlined and annotated.

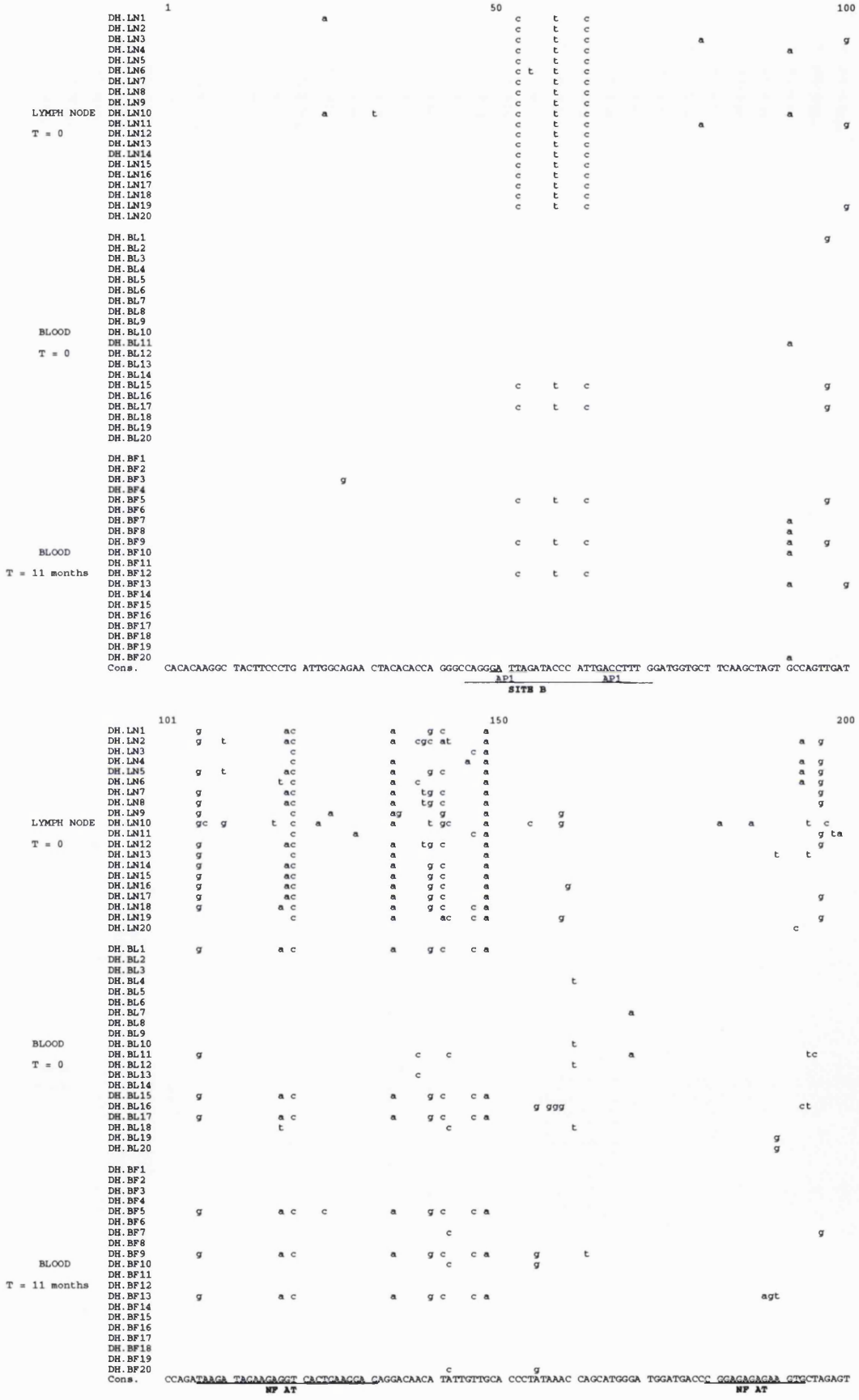

1
50
100
DH.LN1
DH.LN2
DH.LN3
DH.LN4
DH.LN5
DH.LN6
DH.LN7
DH.LN8
DH.LN9
LYMPH NODE DH.LN10
DH.LN11
T = 0 DH.LN12
DH.LN13
DH.LN14
DH.LN15
DH.LN16
DH.LN17
DH.LN18
DH.LN19
DH.LN20
DH.BL1
DH.BL2
DH.BL3
DH.BL4
DH.BL5
DH.BL6
DH.BL7
DH.BL8
DH.BL9
BLOOD DH.BL10
DH.BL11
T = 0 DH.BL12
DH.BL13
DH.BL14
DH.BL15
DH.BL16
DH.BL17
DH.BL18
DH.BL19
DH.BL20
DH.BF1
DH.BF2
DH.BF3
DH.BF4
DH.BF5
DH.BF6
DH.BF7
DH.BF8
DH.BF9
BLOOD DH.BF10
DH.BF11
T = 11 months DH.BF12
DH.BF13
DH.BF14
DH.BF15
DH.BF16
DH.BF17
DH.BF18
DH.BF19
DH.BF20
Cons. CACACAAGGC TACTTCCCTG ATTGGCAGAA CTACACACCA GGGCCAGGGA TTAGATACCC ATTGACCTTT GGATGGTGCT TCAAGCTAGT GCCAGTTGAT
AP1
AP1
SITE B
101
150
200
DH.LN1
DH.LN2
DH.LN3
DH.LN4
DH.LN5
DH.LN6
DH.LN7
DH.LN8
DH.LN9
LYMPH NODE DH.LN10
DH.LN11
T = 0 DH.LN12
DH.LN13
DH.LN14
DH.LN15
DH.LN16
DH.LN17
DH.LN18
DH.LN19
DH.LN20
DH.BL1
DH.BL2
DH.BL3
DH.BL4
DH.BL5
DH.BL6
DH.BL7
DH.BL8
DH.BL9
BLOOD DH.BL10
DH.BL11
T = 0 DH.BL12
DH.BL13
DH.BL14
DH.BL15
DH.BL16
DH.BL17
DH.BL18
DH.BL19
DH.BL20
DH.BF1
DH.BF2
DH.BF3
DH.BF4
DH.BF5
DH.BF6
DH.BF7
DH.BF8
DH.BF9
BLOOD DH.BF10
DH.BF11
T = 11 months DH.BF12
DH.BF13
DH.BF14
DH.BF15
DH.BF16
DH.BF17
DH.BF18
DH.BF19
DH.BF20
Cons. CCAGATAAGA TAGAAGAGGT CACTGAAGCA GAGGACAACA TATTGTTGCA CCCTATAAAC CAGCATGGGA TGGATGACCC GGAGAGAGAA GTGCTAGAGT
NF AT
NF AT

Figure 5.4

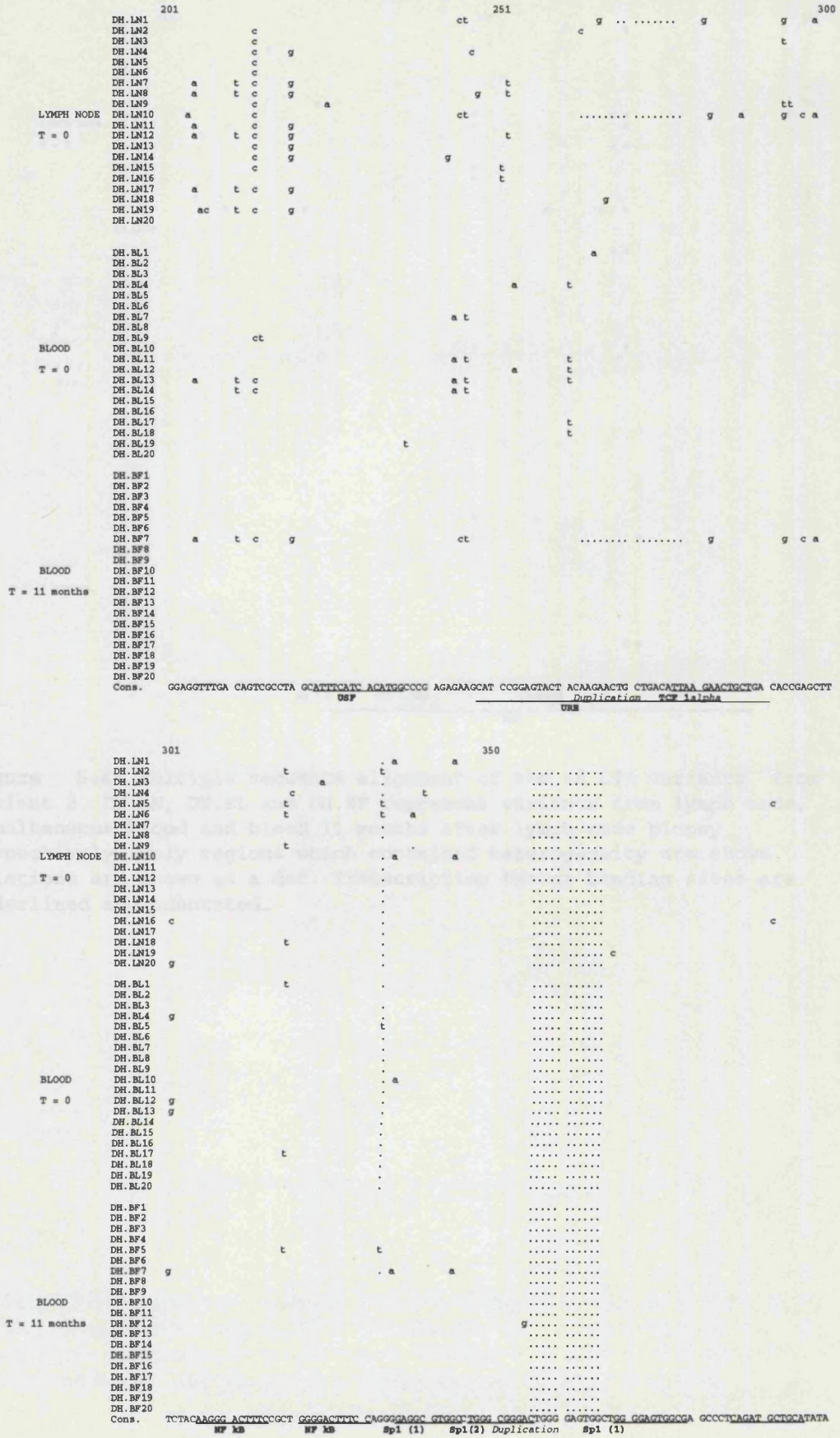


Figure 5.4 (cont)

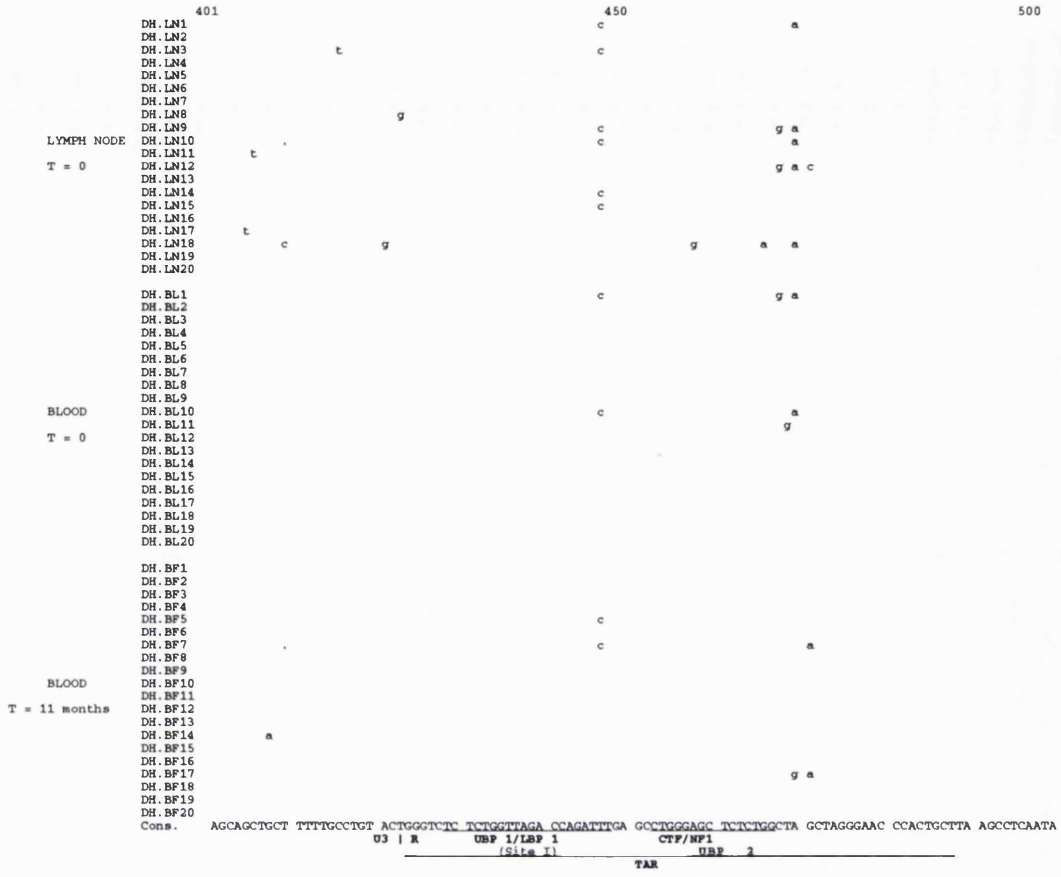


Figure 5.4: Multiple sequence alignment of the 60 LTR variants from patient 2. DH.LN, DH.BL and DH.BF represent variants from lymph node, simultaneous blood and blood 11 months after lymph node biopsy respectively. Only regions which contained heterogeneity are shown. Deletions are shown as a dot. Transcription factor binding sites are underlined and annotated.

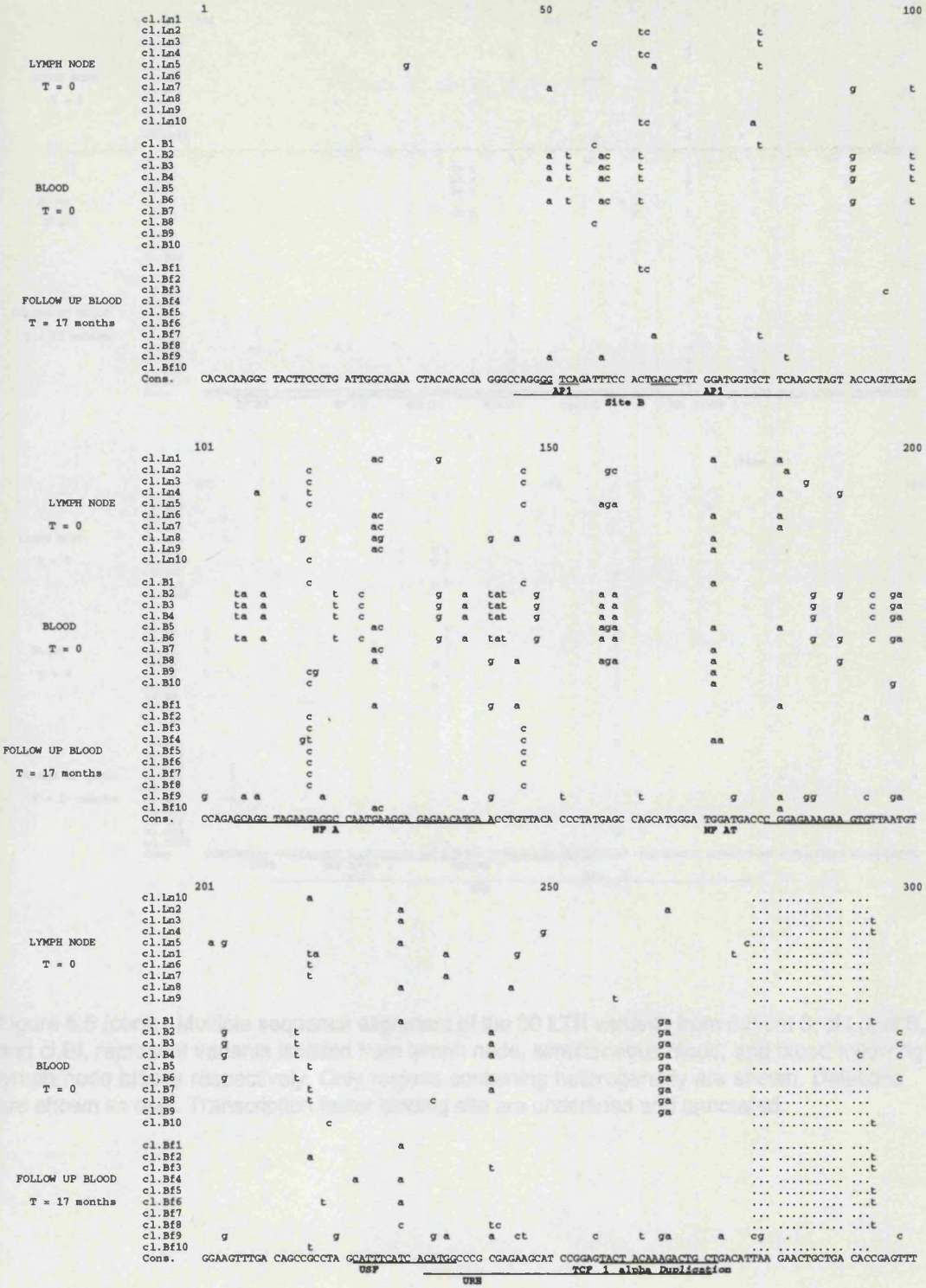


Figure 5.5: Multiple sequence alignment of the 30 LTR variants from patient 3; cl.Ln, cl.B, and cl.Bf, represent variants isolated from lymph node, simultaneous blood, and blood following lymph node biopsy respectively. Only regions containing heterogeneity are shown.Deletions are shown as dots. Transcription factor binding site are underlined and annotated.

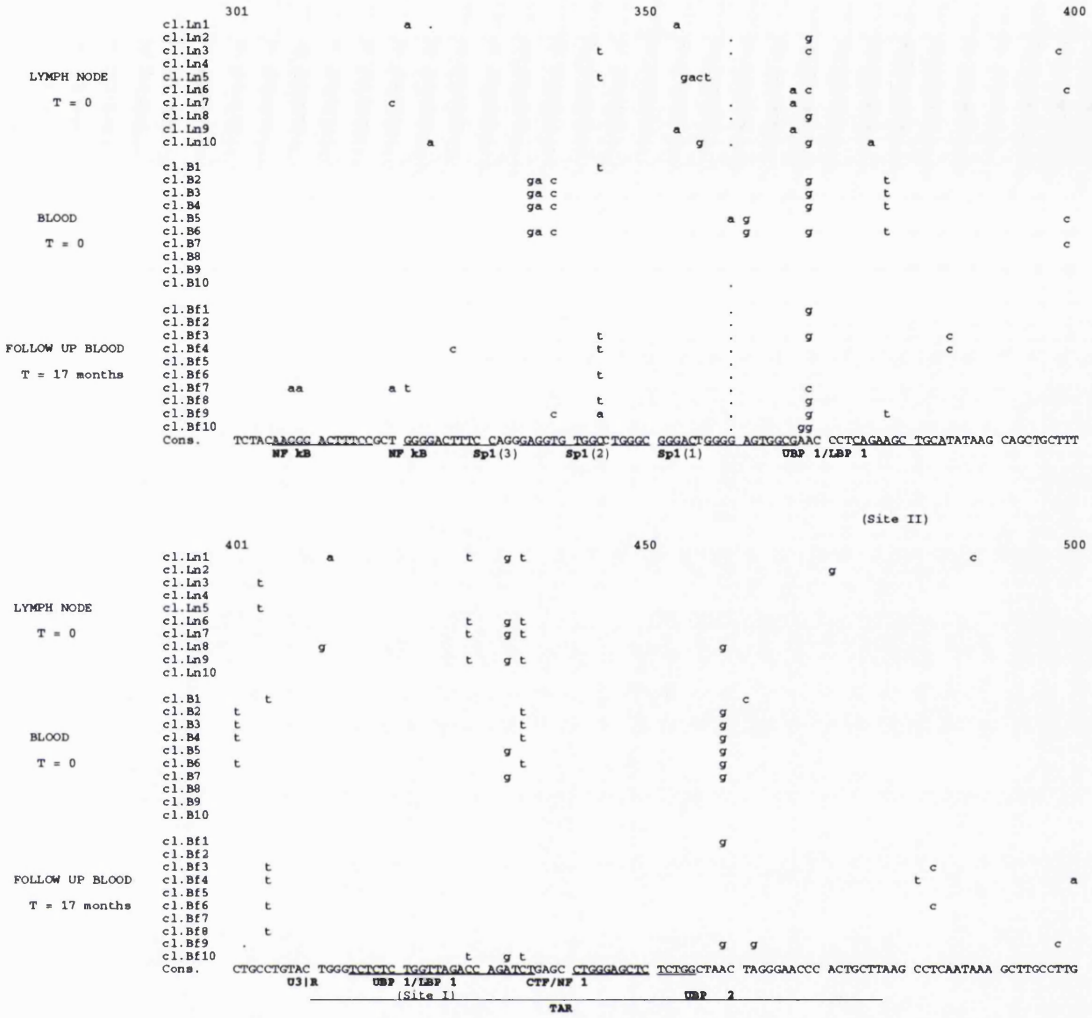


Figure 5.5 (cont) : Multiple sequence alignment of the 30 LTR variants from patient 3; cl.Ln, cl.B, and cl.Bf, represent variants isolated from lymph node, simultaneous blood, and blood following lymph node biopsy respectively. Only regions containing heterogeneity are shown. Deletions are shown as dots. Transcription factor binding site are underlined and annotated.

(a) 5'-AGGGCCAGGGATTAGATACCCATTGACCTTTGGATGGT

5' protein 9 bp spacer 3' protein

binding site binding site

(b) 5'-AGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGT

5' protein 9 bp spacer 3' protein

binding site binding site

Figure 5.6: Schematic representation of HXB2 Site B (a) and the natural variant from patient 2 (b). The 5' and 3' protein binding sites, and the 9 bp spacer are labelled.

5.2.3 Band shift analysis on a Site B natural variant

DNA mobility shift assays were performed on the Site B variant detected in patient 2 LTR quasispecies. Competitive binding experiments were carried out using a 30bp Site B wild type ^{32}P radiolabelled oligonucleotide probe which sequence is 5' AGGGCCAGGGATTAGATACCCATTGACCTTT, and either a cold (non-labelled) wild type Site B oligonucleotide (as above) or a cold Site B variant oligonucleotide: 5' AGGGCCAGGGAT**C**AGATA**T**CCA**C**TGACCTTT (see figure 5.6).

The amount of non-radiolabelled competitor was increased from zero (absent) to 2000 fold in excess. As shown in figure 5.7, the nucleoprotein/DNA complex is strongest in lane 1 where no competition was applied, then decreases in intensity as more competitor is added. Close to complete competition is achieved with X 100 fold excess of wild type competitor (lane 5'), whereas a 500-fold excess of Site B variant competitor must be present to achieve a similar extent of competition (lane 6). Cerenkov counting was applied to the radioactive polyacrylamide gel portions corresponding to the shifted probe and the free probe (bottom of the gel), and the percentage of probe shifted (%SHIFTED) calculated as the total amount of probe divided by the amount of probe shifted. The results were then expressed as a histogram (Figure 5.8). The resulting Histogram showed that while a 10 fold excess of wild type competitor is required to decrease the % of probe shifted to approximately 5%, a 100 fold excess of Site B variant competitor is required to achieve a similar decrease in amount of probe shifted. This DNA mobility shift assay showed that the protein binding factor has a 10 fold less affinity for the natural site B variant than for the HXB2 prototype Site B.

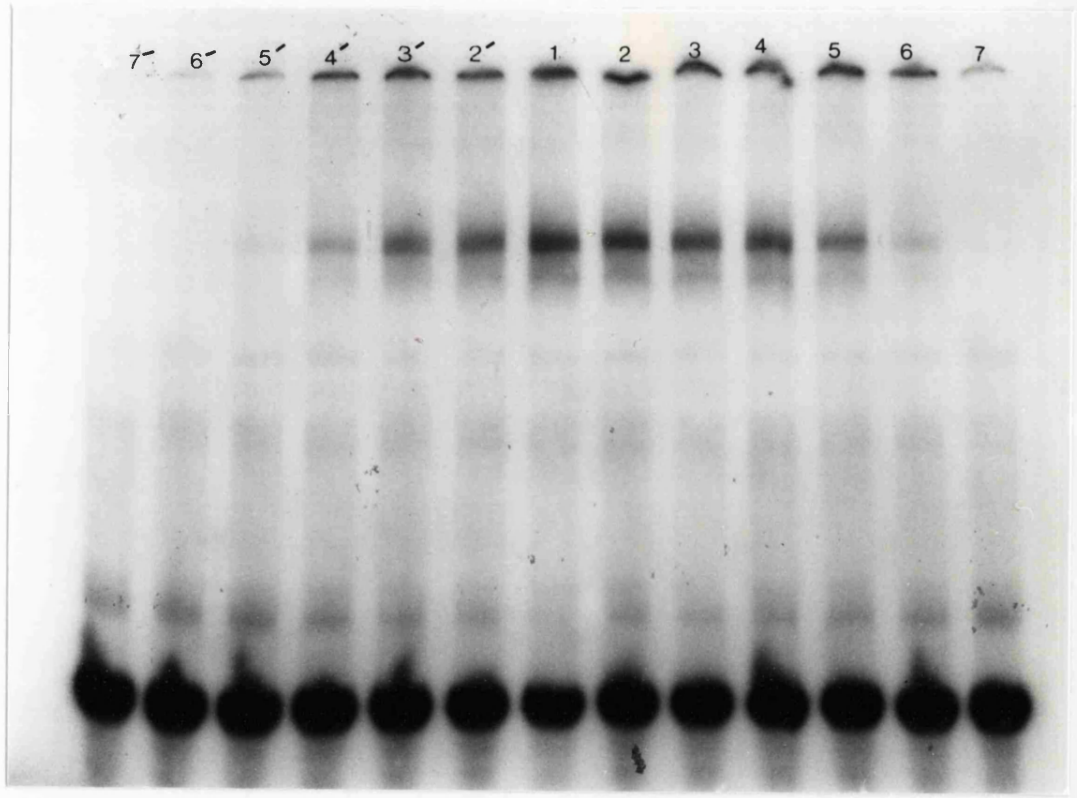


Figure 5.7: Autoradiograph of a competive band shift assay. The amounts of competitor used were zero, 5, 10, 20, 100, 500, 2000 fold excess corresponding to lane1 (X0), then to lanes 2, 3, 4, 5, 6, and 7 respectively for Site B variant ; and to lanes 1', 2', 3', 4', 5', 6', and 7' for wild type Site B.

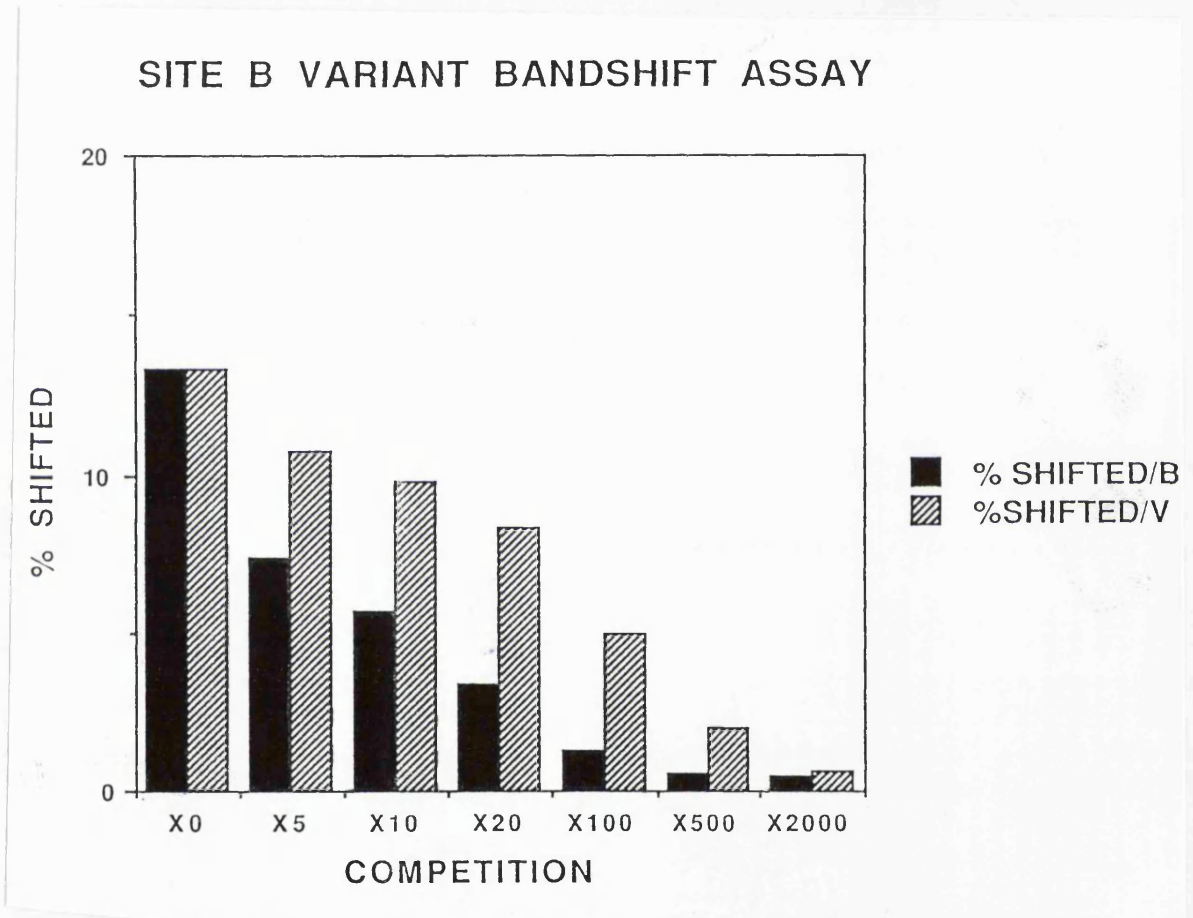


Figure 5.8: Histogram of Site B variant band shift assay. The percentage of probe shifted when competed with wild type is represented in plain black bars, and the percentage of probe shifted when competed with site B variant is represented as dashed grey bars.

5.2.4 Divergence of the LTR quasispecies

The multiple sequence alignments were used via PHYLIP DNADIST output as the basis for the calculation of the genetic distance between variants in each compartment for each patient (Tables 5.1, 5.2, and 5.3).

For patient 1, there was no significant differences in the heterogeneity between the lymph node and simultaneous blood quasispecies, which were 1.31% and 1.36% respectively. This patient had a relatively high CD4 cell count (220 x 10^6/l) at the time of biopsy, had CDC stage IV disease and had extensive disruption of lymph node architecture. There was a statistically significant difference in heterogeneity between the blood quasispecies corresponding to the samples obtained 14 months prior to biopsy and 6 months after the biopsy, divergence at both time points were 1.69% (Table 5.1), i.e., greater than divergences of both LN and simultaneous blood.

Patient 2, harboured a lymph node HIV-1 quasispecies that was more heterogeneous (3%) than either the simultaneous blood quasispecies (1.59%) or the 11 month follow-up blood quasispecies (1.38%). This patient had a relatively low CD4 cell count (18 x 10^6/l) at the time of biopsy, was classified as having CDC stage IV disease, but interestingly, had a histologically intact lymph node architecture at the time of biopsy (see table 5.2).

In the case of patient 3, there was no significant difference in the heterogeneity between the lymph node (3.62%) and simultaneous blood quasispecies (5.66%). This patient had a relatively high CD4 cell count at biopsy (680 x 10^6/l), had CDC stage III disease, but histology revealed extensive destruction of the lymph node

architecture. The quasispecies isolated from the follow up blood sample (+17 months) when the CD4 count had declined to 420x 10^6/l, had a lower median divergence of 2.46% (see table 5.3).

5.2.5 HIV-1 LTR phylogenetic analyses

In order to further understand the intra and inter-relationships between individual variants within each patient sample DNA parsimony phylogenetic methods with bootstrap resampling have been applied to the data set. The resulting consensus phylogenetic trees are shown for patients 2, 1, and 3 in Figures 5.9, 5.10, and 5.11 respectively. In patient 2 (Fig. 5.9) the consensus tree showed a clear dichotomy between the lymph node and peripheral blood variants. This resulted in a polarised tree with two major branches: 32 out of 40 blood variants were contained in one branch (peripheral blood variants clade), whilst 19 out 20 lymph node variants comprised the other branch (lymph node variants clade) . This polarisation was detected in greater than 90% of the bootstrap resamplings.

TABLE 5.1: MEDIAN DIVERGENCE BETWEEN THE LTR VARIANTS PRESENT IN THE LYMPH NODE AND BLOOD SAMPLES OF PATIENT 1

Time (months)	Absolute CD4 Count x 10^6/L	Histology	Median Divergence	
			Lymph Node	PBMC
-14	270	-	NA	0.0169
+0	220	Follicular lysis dissruption of germinal centres	0.0131	0.0136
+6	110	-	NA	0.0169

Lymph node/simultaneous blood: p = NS; Lymph node/blood before biopsy: p = 0.02;
Lymph node/follow-up blood: p = 0.0008.

TABLE 5.2: MEDIAN DIVERGENCE BETWEEN THE LTR VARIANTS PRESENT IN THE LYMPH NODE AND BLOOD SAMPLES OF PATIENT 2

Time (months)	Absolute CD4 Count x 10^6/L	Histology	Median Divergence	
			Lymph Node	PBMC
0	18	Intact Germinal Centres	0.03	0.0159
+11	7	-	NA	0.0138

Lymph node/simultaneous blood: $p < 0.0001$
Lymph node/follow-up blood: $p = 0.0001$

TABLE 5.3: MEDIAN DIVERGENCE BETWEEN THE LTR VARIANTS PRESENT IN THE LYMPH NODE AND BLOOD SAMPLES OF PATIENT 3

TIME (MONTHS)	ABSOLUTE CD4 COUNT X 10^6/L	HISTOLOGY	MEDIAN DIVERGENCE	
			Lymph Node	PBMC
0	680	Extensive destruction & granulomas	0.0362	0.0566
+17	420	NA	NA	0.0246

Lymph node/follow-up blood: p = 0.002

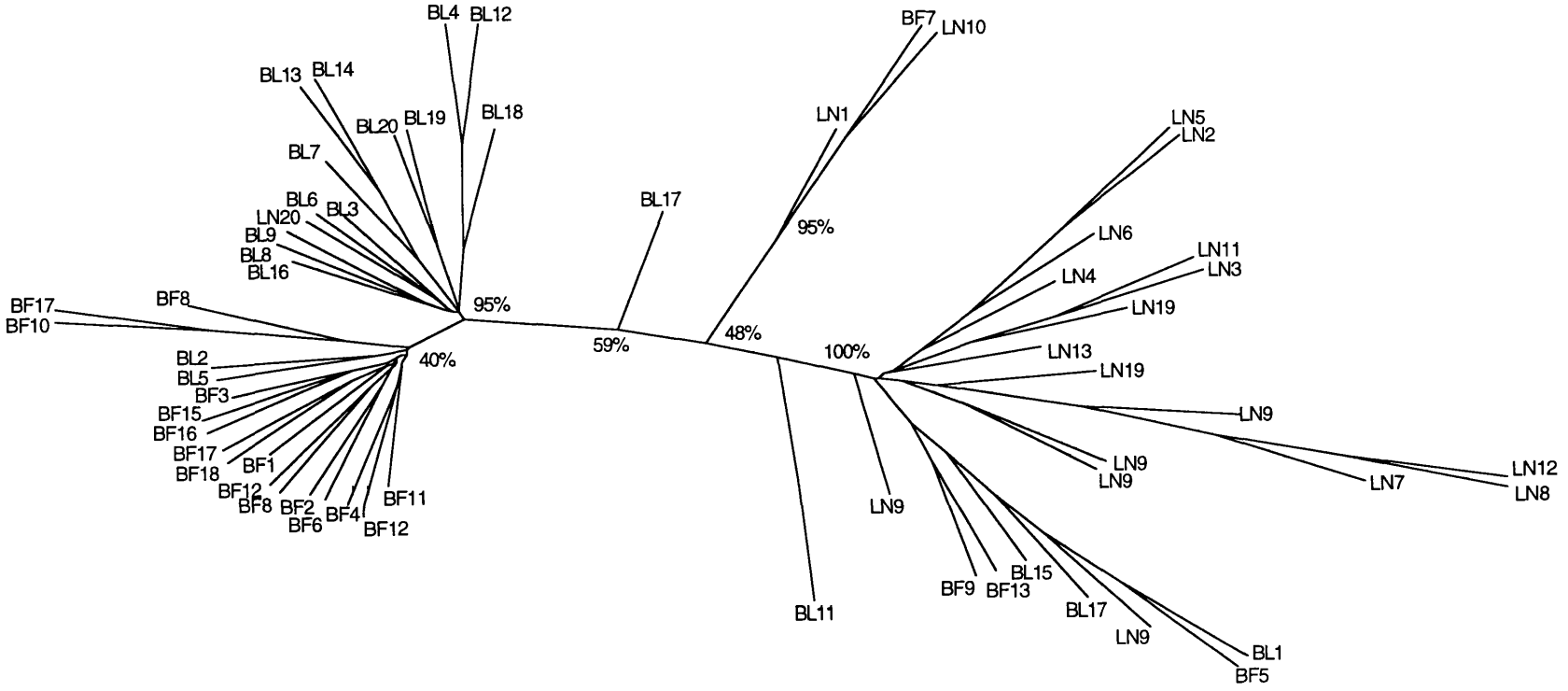


Figure 5.9: Maximum parsimony tree of LTR variants from patient 2. Variants were obtained from the lymph node (LN), simultaneous blood sample (BL), and 11 month follow-up blood sample (BF). All branches are drawn to scale, and relevant bottstrap values are annoted.

In contrast, the consensus trees constructed for patients 1 and 3 (Fig. 5.10 and 5.11 respectively) are substantially different from that of patient 2 and demonstrate no evidence of polarisation between lymph node and peripheral blood variants. Indeed the bootstrap values of the branches are very low, indicative of very closely related variants. Moreover, lymph node and blood variants are present in common branches. Interestingly, in the case of patient 1, LTR variants from the blood sample obtained 14 months prior to the lymph node biopsy, clustered together, independently of the lymph node variants, whereas the blood variants from the blood obtained simultaneously to biopsy and follow up blood were closely intermingled with the lymph node variants. This result suggest that the compartmentalisation between lymph node and blood variant was compromised by the disrupted histological structure of the lymph node at the time of biopsy, and that its structure might have been less damaged 14 month prior to histological analysis, when the "blood before" sample was obtained.

These phylogenetic trees are consistent with the data on genetic heterogeneity shown in Tables 5.1, 5.2, and 5.3. The phylogenetic trees have also been constructed using alternative phylogenetic approaches such as Fitch-Margoliash and the polarised topology observed between lymph node and peripheral blood variants in patient 2 remained a consistent feature (data not shown).

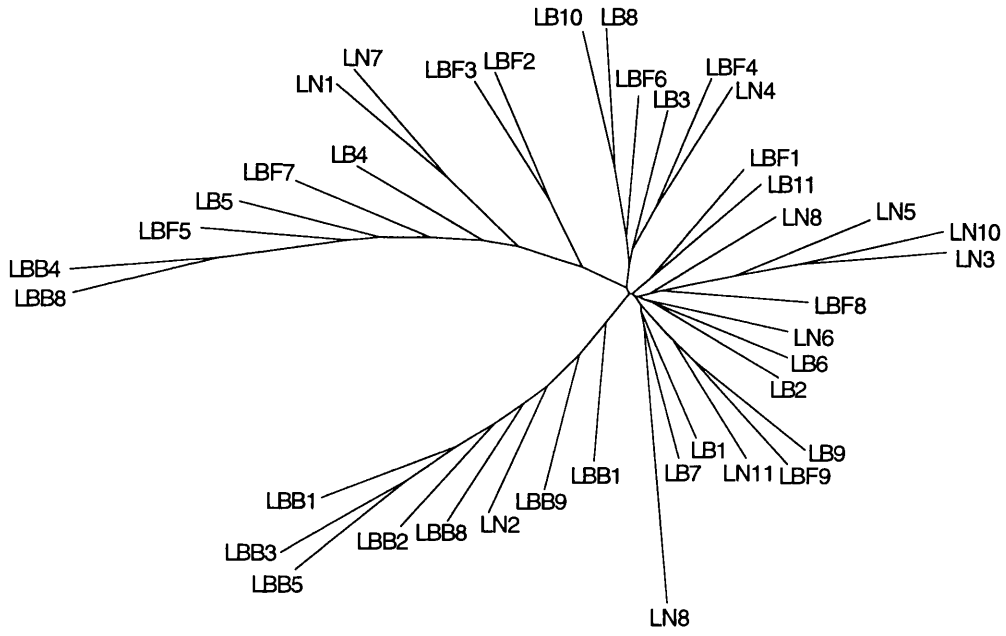


Figure 5.10: Maximum parsimony tree of LTR variants from patient 1. LN, LB, LBB, and LBF, represents variants from lymph node, simultaneous blood, blood before biopsy, and blood after biopsy respectively. All branches are drawn to scale, and all bootstrap values were low.

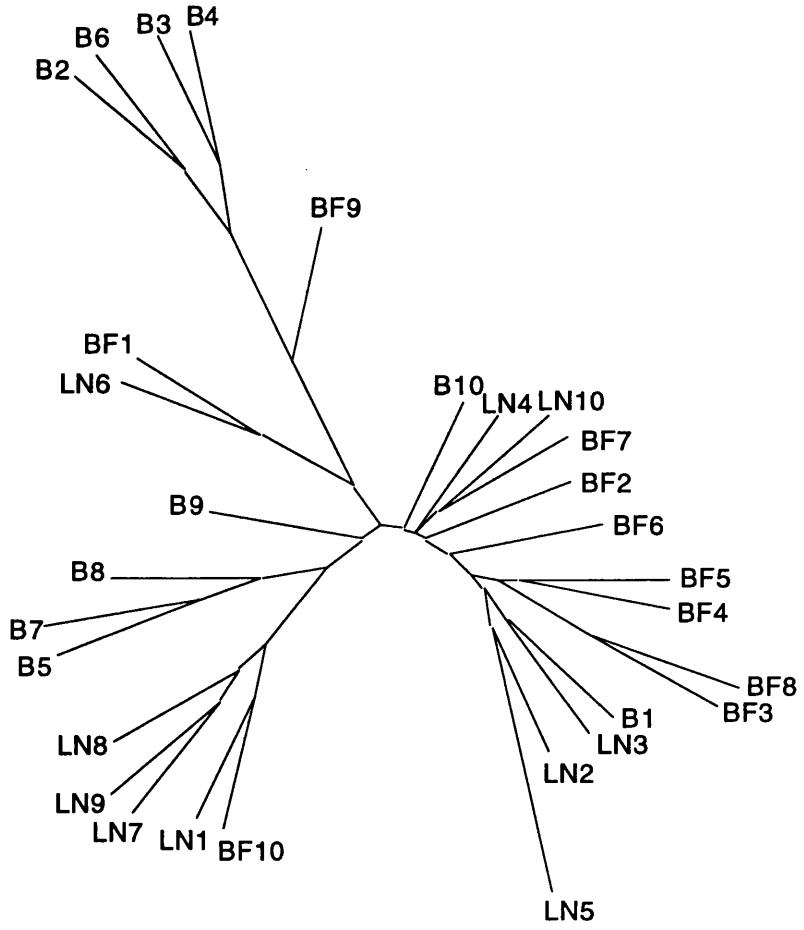


Figure 5.11: Maximum Parsimony tree of LTR variants from patient 3. LN, B, and BF, represent variants from lymph node, simultaneous blood sample, and blood sample following biopsy respectively. All branches are drawn to scale, and all bootstrap values were low.

5.3 DISCUSSION

The HIV-1 LTR variants present in the lymph node and peripheral blood of three HIV-1 infected individuals have been analysed. The multiple sequence alignment analyses allow several conclusions to be drawn. Firstly, that the NF-κB sites, Sp1 sites, and TAR element are very well conserved regions of the LTR, indeed more than 90% of the natural variants possess a prototypic sequence for these transcription binding sites. Only one variant from patient 2 possesses a duplication generating a fourth Sp1 site and none of the 140 natural variants described possessed an NF-κB duplication. These results show that these regions have a crucial function in HIV-1 transcription and confirm the functional data reported in other studies (Roy et al., 1990; Alcami et al., 1995; Harrich et al., 1995). Secondly, in my study, the length polymorphism of the LTR region is almost exclusively due to a duplication of the TCF-1α region. Indeed similar duplications have been reported in several other studies (Golub et al., 1990; Englund et al., 1991; Koken et al., 1992; Michael et al., 1994), these duplications conferred different effects on the LTR function depending on the presence or absence of other mutations in the LTR (Koken et al., 1992; Michael et al., 1994). These studies have shown that mutations and duplications have complementary effects, and therefore, that no mutation can be taken in isolation for functional studies. In this respect, DNA mobility shift assays looking at particular regions in isolation should be interpreted with caution. Thirdly, the negative regulatory element (NRE) of the LTR (-420 to -159), comprising the site B region, the 2 NF-AT sites, and USF region, show greater heterogeneity than the enhancer-promoter region (-158 to -1) or the TAR element (+1 to +57). The higher heterogeneity of the NRE demonstrate the sequence flexibility of this region (under less selective pressure), and suggests a lesser importance of this region for the LTR transcriptional potency.

Finally a variant of the NRE site B was identified, comprising 3 mutations; a T → C at position -361 (inside the 5' protein binding site), C → T at position -355 and T → C at position -351 (inside the 9 bp spacer). DNA mobility shift assays showed that the site B recognition protein possesses a 10 fold lower affinity for this variant than for the wild type sequence. This decreased affinity may suggest a less potent NRE and therefore a higher transcriptional efficiency of these variants. However, as already mentioned, DNA mobility shift assays looking at a particular region in isolation should be interpreted with caution.

The genetic variation of the LTR quasispecies present in the lymph node was significantly higher than that present in the peripheral blood quasispecies in the patient who possessed a lymph node with an intact follicular dendritic cell network and germinal centres. In the two patients in which substantial disruption of the lymph node structure had occurred, including follicular lysis and absence of germinal centres, there was no difference in genetic diversity between the two compartments. The higher heterogeneity of the lymph node LTR quasispecies of patient 2 suggests that intact lymph nodes may sustain active replication (required for the generation of diversity) and provide a niche in which HIV strain evolution can occur more rapidly and be sustained . The results of a study comparing the phenotypes of lymph node and blood variants isolated at simultaneous times from 11 HIV-1 infected individuals, are in accordance with this hypothesis (Tamalet et al., 1994). The results of Tamalet and collaborators showed that viral burden was 1.73 log higher in lymph node than blood in patients with intact lymph nodes and only 0.37 log higher in patients with AIDS-related complex; moreover 5/11 lymph node bulk isolates were phenotypically distinct (NSI versus SI) from autologous PBMC isolates, finally, in three patients, the autologous serum neutralised the PBMC isolates but not the Lymph node isolates.

These result suggest that the relatively high level of HIV-1 replication in lymph nodes may favour the emergence of variant viruses exhibiting specific phenotypes. Lymph nodes have been shown to contain a higher HIV burden than PBMCs (Pantaleo et al., 1993a; Embretson et al., 1993; Tamalet et al., 1994; Lafeuillade et al., 1995). Although we have not determined which cells within the lymph node gave rise to the HIV quasispecies analysed in this study, results of other studies indicate that the majority of cells harbouring HIV proviral DNA are CD4+ lymphocytes and macrophages and not follicular dendritic cells (Embretson et al., 1993; Donaldson et al., 1994a). Follicular dendritic cells appear to trap virions on their villus processes early in the course of HIV infection and may transmit infection to cells as they migrate through the lymph node. However, further data from many sources must be obtained on more individuals before the dynamics of strain evolution in multiple tissue sites and the crossover of strains between these sites is fully understood.

Phylogenetic analysis indicated that blood and lymph variants independently clustered together in clearly separated clades in patient 2. Moreover, 3 blood variants obtained at the time of lymph node biopsy, 3 variants from the 11 months follow up blood sample and 1 lymph node variant clustered together, forming a sub-branch in the major lymph node clade. These peripheral blood variants are therefore closely related to the lymph node variants. In contrast to the results obtained for patient 2, a distinct clustering of variants was not observed in patients 1 and 3 whose lymph nodes exhibited extensive histological deterioration. The short inter-node branches observed in the consensus trees for these patients indicates that the LTR variants analysed are closely related to each other with lymph node and blood variants present in common branches. The polarised topology of the consensus tree obtained for patient 2 with a significant branch-length separating the lymph node

and peripheral blood clades provides data in support of the hypothesis that independent evolution of HIV-1 variants can occur in these two compartments when the lymph node is intact.

The results presented in this study, albeit in small numbers of patients, suggests that intact germinal centres and follicular dendritic cell network within the lymph nodes may be an important factor in the sequestration/evolution of distinct HIV variants. Hence, disruption of the lymph node histological structure would lead to seeding of HIV variants in the periphery; a prediction that is consistent with the data obtained from patients 1 and 3. However, in order to provide sufficient data for statistically significant conclusions of this type to be drawn, similar studies on greater numbers of patients are required. It should be noted that these analyses are confounded by the fact that lymph node biopsies are not routinely taken from all HIV infected patients, especially in the early stages of disease. Such sampling problems could be overcome by performing similar analysis in appropriate animal models of HIV e.g. SIV infected macaques.

Furthermore, the functional effects of sequence variations within the LTR present in the patients analysed and the subsequent disease course should be investigated. In this context, it is interesting to note that the clinical status of patient 2 remains relatively stable despite having low CD4 cell numbers and CDC stage IV disease, whilst patient 1 has had a continuously declining CD4 cell count and had progressed rapidly to CDC stage IV disease.

Further data on more patients must be accrued before the precise role of the lymph nodes in HIV strain evolution and pathogenesis is defined. However, the data

presented in this chapter, illustrate the importance of performing both cross-sectional and longitudinal quasispecies analyses on many target organs within HIV infected individuals in order to understand the complex interplay between the HIV and its host.

CHAPTER 6

DISTRIBUTION OF HIV-1 LTR VARIANTS IN MULTIPLE BODY SITES: DIFFERENCES BETWEEN HIV-1 LONG TERMINAL REPEAT QUASISPECIES PRESENT IN NERVOUS TISSUES AND THAT IN LUNG, BLOOD AND LYMPHOID TISSUES OF AN AIDS PATIENT.

6.1 INTRODUCTION

The coexistence of genetically distinct HIV-1 natural variants within individuals has been clearly demonstrated (Goodenow et al., 1989; Meyerhans et al., 1989; Simmonds et al., 1990; Balfe et al., 1990; Delassus et al., 1991) but the role of these multiple variants in the pathogenesis of AIDS is still unclear. To date, it has been reported that the V3/V4 region of the HIV-1 env gene in brain and spleen exhibit a tissue-specific but host-dependent evolution (Epstein et al., 1991; Keys et al., 1993), whilst other studies on both the env region and the long terminal repeat (LTR) have shown that strain evolution in infected organs and peripheral blood can also proceed in parallel (Pang et al., 1991; Delassus et al., 1992a; Ait-Khaled and Emery, 1994; Ball et al., 1994). Investigations on the cellular tropism of HIV have mainly concentrated on the env gene and specifically the V3 region; such studies have demonstrated that regions outside the CD4 binding site and comprising the V3 loop, are critical for binding to and/or entry into human brain microglia (Cheng-Mayer et al., 1990; Collman et al., 1992; Fouchier et al., 1992; Kuiken et al., 1992; Westervelt et al., 1992; Takeuchi et al., 1991; Sharpless et al., 1992; Korber et al., 1992). In addition, a study based on a cohort of 22 HIV-1 infected patients from which brain biopsies were obtained suggested that specific amino acid changes in the V3 region (positions 305 and 329) were associated with AIDS dementia (Power et al., 1994). Power and colleagues extended the finding of other studies by showing that V3 sequences from brain possess the genetic determinants for a macrophage tropic phenotype, and present data suggesting that distinct HIV envelope sequences are associated with the clinical expression of HIV dementia. However, the specific amino acid changes reported in the above study, have not been reported in other cohorts of HIV infected patients with dementia and therefore remains to be confirmed.

As summarised in chapter 1, the involvment of the V3 region in tissue tropism, does not exclude other regions of the HIV genome from being important in HIV adaptation to different tissues after viral entry. The LTR region of HIV is also likely to have a role in HIV adaptation to different tissues subsequent to initial infection since it contains multiple functional elements required for the differential and temporal expression of the proviral genome (Gaynor, 1992). The LTR region of the retroviruses has been implicated in tissue-specific gene expression and disease for Rous sarcoma virus, Visna virus, murine leukemia virus and HIV (Lenz et al., 1984; Rosen et al., 1985a; Holland et al., 1985; Overbeek et al., 1986; Small et al., 1989; Koyanagi et al., 1987; Nakanishi et al., 1991; Moses et al., 1994).

Neuropathogenic strains of HIV possessing distinct biological, serological and molecular properties from peripheral blood isolates have also been identified (Anand et al., 1987; Cheng-Mayer et al., 1989; Watkins et al., 1990). Indeed, the LTR of a natural variant of HIV-1 isolated upon culture of the cerebrospinal fluid of an AIDS patient ,HIV-1 JR csf, (Koyanagi et al., 1987) has been shown to direct CNS specific expression in transgenic mice(Corboy et al., 1992). In contrast, transgenic mice containing an LTR-β galactosidase construct based on the T-cell tropic strain of HIV-1 (HIV-1 IIIB/LAV), directed expression in the thymus but not in the CNS of the mice (Corboy et al., 1992). These investigators have used the aforementioned transgenic mice as well as transgenic mice expressing β galactosidase under the control of the LTR of JRfl (an isolate from the frontal lobe of patient JR) to examine the effects of neurodevelopmental changes in the expression of HIV-1 in the CNS (Buzy et al., 1995). Analysis of the transgenic mice at embyonic to early postnatal stages showed that the CNS-derived HIV-1 LTRs were capable of regional, temporal, and cellular specificity in their ability to direct gene expression. Overall, CNS expression

directed by these LTRs was higher at early postnatal times than in the adult transgenic mice. These data suggested that cellular transcription factors normally present during CNS development contributed to increased viral gene expression. It is important to note that in both the above studies (Corboy et al., 1992; Buzy et al., 1995), the major CNS infected cells were neurones. This would not be the case in natural HIV-1 infection, and is explained by the absence of requirement for an interaction between the virus and its cellular receptor which is indeed bypassed in the transgenic mouse. Thereby, transgenic mice allow the specific mechanism of tissue-specific transcriptional regulation and viral gene expression to be investigated.

Moreover, TAR independent *tat* activation of replication in the central nervous system through a CNS specific κB binding activity has also been reported (Taylor et al., 1992; Bagasra et al., 1992; Taylor et al., 1994). Thus, some strains of HIV-1 appear to possess a selective advantage for gene expression in nervous tissues by responding to the specific cellular transcription factors preferentially present in these tissues (Corboy et al., 1992; Taylor et al., 1994). At present, the microglial or more generally, the nervous tissue specific transcription factors that interact with HIV-1 LTR are ill-defined. A recent study showed that a glial cell line was more permissive to HIV-1 replication than a neuroblastic cell line (Ensoli et al., 1994). This cell type-specific replication was conferred by the LTR as demonstrated by LTR-CAT expression in the two cell lines. Moreover, the study demonstrated the ability of nerve growth factor (NGF) to up-regulate HIV-1 gene expression in glial cells but to a lesser extent in neuronal cell lines, suggesting the low expression of complementary positive regulatory factor(s) in those latter cells (Ensoli et al., 1994).

Finally, Taylor and colleagues have shown that CNS-derived cells express κB binding factors that differ from prototypical NF-κB. They also demonstrated that this CNS-specific κB-binding activity enhances HIV-1 transcription *in vitro* and facilitates TAR-independent *tat* transactivation (Taylor et al., 1994). Thus, in addition to tropism dictated by *env* sequences, some strains of HIV-1 appear to possess a selective advantage for gene expression in the CNS by responding to the specific cellular transcription factors present in this tissue.

In the light of the aforementioned data implicating the LTR of retroviruses in tissue specific expression, I investigated the sequence variation of HIV-1 LTR variants present in post-mortem samples of lymph node (LN), dorsal root ganglion (G), lung (LU), spinal cord (SC) and spleen (SP) as well as a pre-mortem peripheral blood (B) sample of a patient dying with AIDS, and related the results to clinico-pathological findings and discuss the role of the HIV LTR as a potential determinant of tissue adaptation, especially in the nervous system.

6.2 SUMMARY OF PATIENT 4 CLINICAL HISTORY AND SAMPLES ANALYSED

6.2.1 Clinical course

The patient first presented in January 1990 with a 4 week history of cough and dyspnoea, weight loss, sweating and rash and was subsequently diagnosed HIV-1 antibody positive. *Pneumocystis carinii* and cytomegalovirus was isolated from bronchial lavage washings. The patient said that the only occasion he had been at risk of acquiring HIV was six weeks before his first admission. The absolute CD4 count at this time was 378 x 10^6/l. At discharge the patient was prescribed zidovudine (200 mg 5 times a day) and nebulised pentamidine (300mg monthly). The patient was subsequently treated for a succession of pathological conditions:

cytomegalovirus (CMV) pneumonitis (1/90) requiring ventilation, cutaneous Kaposi Sarcoma (KS) (9/91), CMV oesophagitis (1/92), pleural effusion (2/92), pulmonary KS and CMV in broncho-alveolar lavage (BAL) (10/92). The patient started an alternating regimen (every 2 months) of zidovudine and didanosine (200mg twice daily) in April 1992 and died in November of the same year.

6.2.2 Processing of samples

A peripheral blood sample (10ml) in preservative free heparin was obtained in September 1992, 2 months prior to death. Full autopsy was undertaken by Dr J. McLaughlin and the following tissues were further investigated: spleen, lung, lymph node, spinal cord and dorsal root ganglion. PBMCs were purified by Ficoll-Hypaque gradients. DNA was extracted and purified from the post-mortem tissue ($1cm^3$ frozen tissue blocks) and PBMCs by proteinase K digestion, phenol extraction and ethanol precipitation using methods described in Chapter 2 (section section 2.3.1). Virus culture was carried out by the diagnostic virology team, at the Royal Free Hospital, UK. In brief, buffy-coat and minced post-mortem samples were inoculated onto primary human embryo lung fibroblasts as previously described (Pillay et al., 1993). Infected cells were maintained for 3 weeks with regular examination for the presence of viral cytopathic effects. Adenovirus presence in the culture supernatants was confirmed by electron microscopy.

6.2.3 Amplification of HIV-1 LTR variants from multiple tissues

The 540 bp LTR amplicon of patient 4 samples was cloned and at least 5 clones from each tissue was sequenced, see Chapter 2 (sections 2.5 to 2.8)). Representative autoradiograph of DNA sequencing gel autoradiographs encompassing the TAR element are shown in figures 6.1, 6.2.

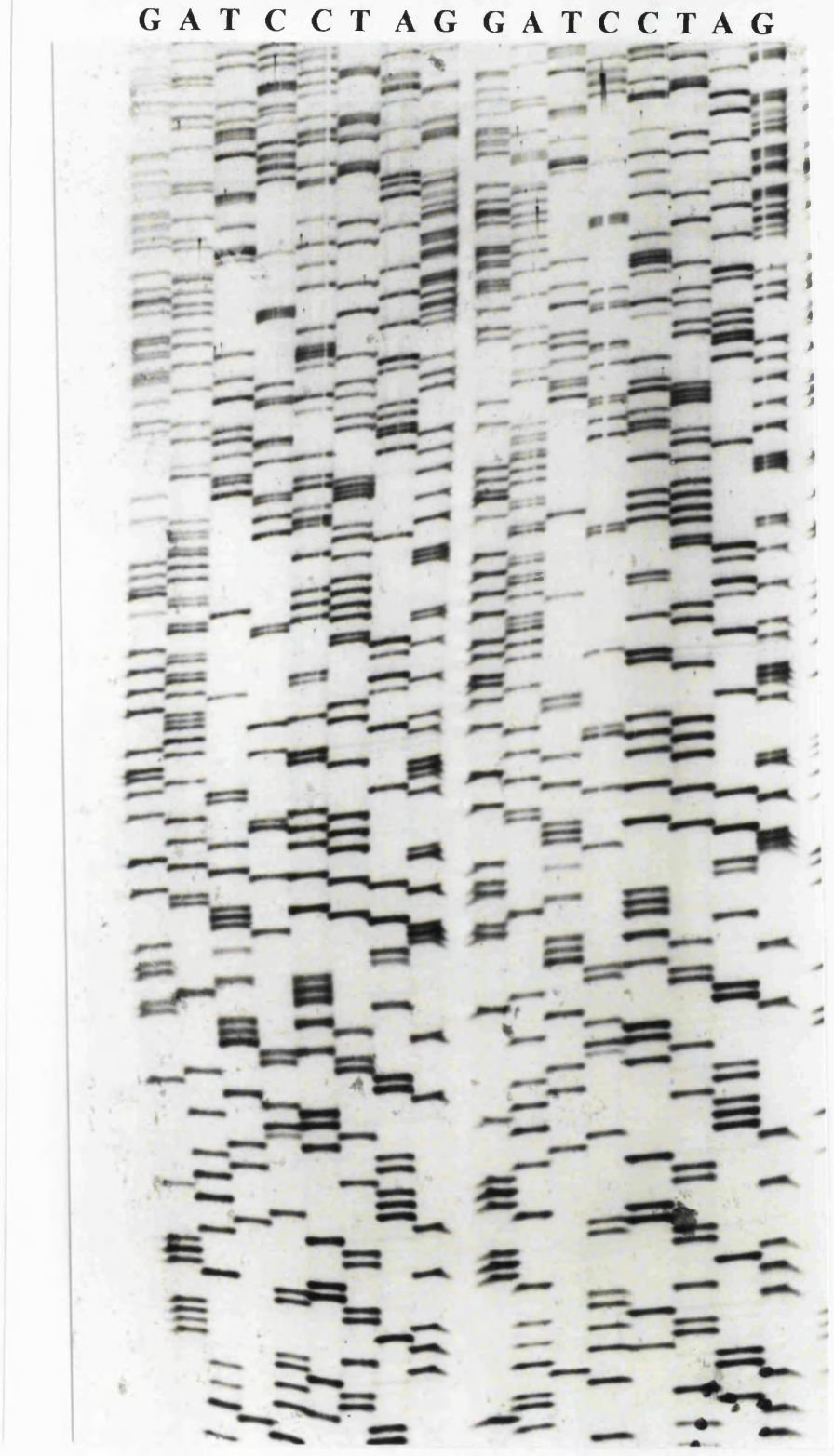


Figure 6.1: Autoradiograph of LTR clones LN1 & 2 from patient 4.

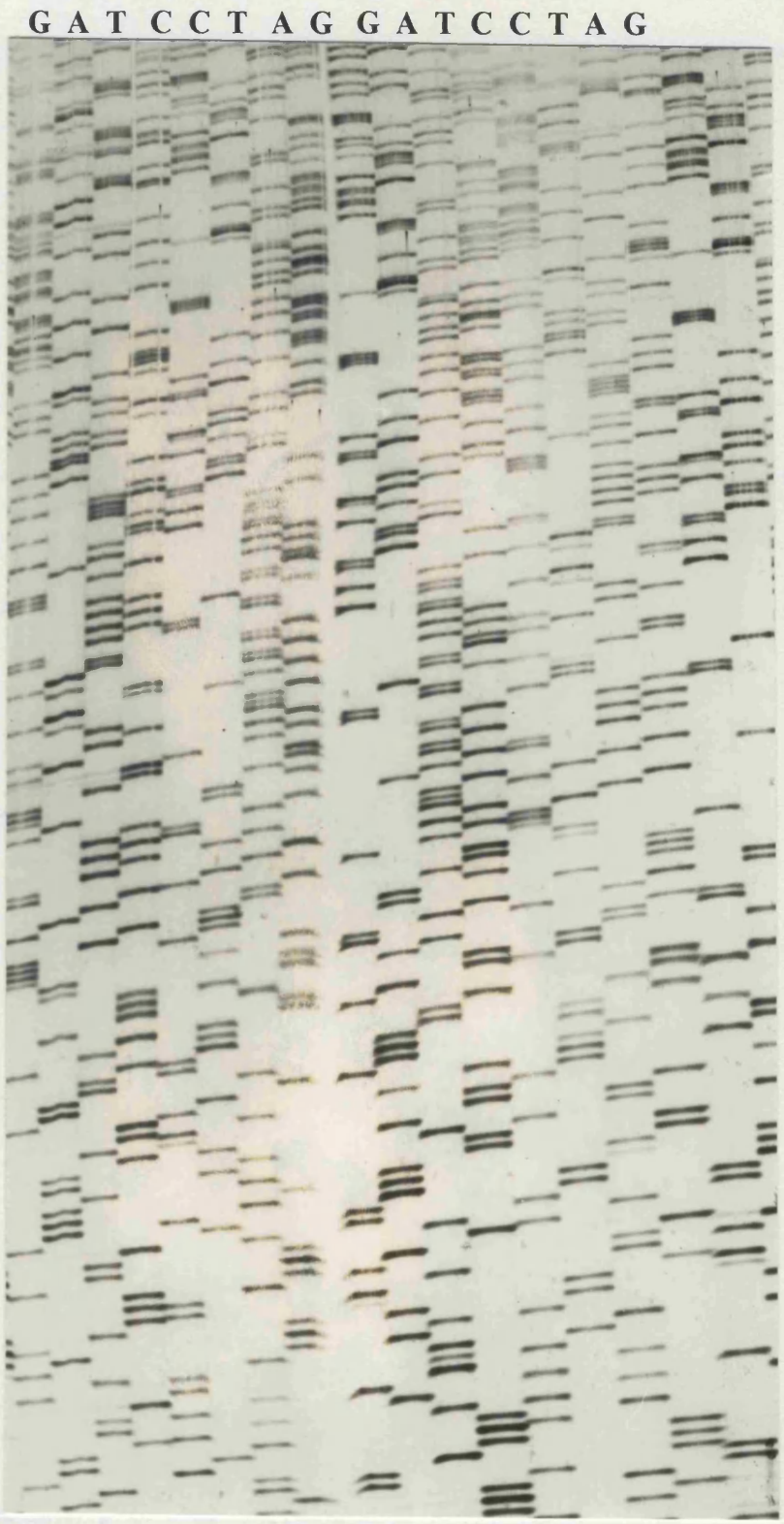


Figure 6.2: Autoradiograph of LTR clones Spleen 4 (left) and LN4 (right) from patient 4.

6.3 CLINICO-PATHOLOGICAL ANALYSIS

The patient described in the current study was a male homosexual dying with AIDS 34 months after presenting at the Royal Free Hopital with probable HIV-1 seroconversion illness and a CD4 count of 378 x 10^6/l. During the 24 months prior to death his absolute CD4 levels fell from 80 x 10^6/l to 10 x 10^6/l at 12 months and remained stable at this level until death. The results of the histopathological and virological analysis of the post-mortem tissues that were subsequently subjected to HIV-LTR specific PCR and phylogenetic analyses are shown in Table 6.1. All tissues analysed showed abnormalities on histology. The lymph node and spleen exhibited lymphoid depletion whilst the lung sample showed evidence of extensive opportunistic infection with *P. carinii* and yeast. The spinal cord showed increased numbers of corpora amylacea at all levels, prominence of microglia with microglial nodules in the anterior horns, and several foci of vacuolar myelopathy. The latter was particularly prominent in the posterior and postero-lateral column at the T9 level and in one anterior column at the L2 level (Figure 6.3). However, the patient had no clinical evidence of HIV-1 dementia prior to death although post mortem examination of the brain revealed low grade lepto-meningitis and multiple microglial nodules in the basal ganglion and brain stem. No multinucleated giant cells characteristic of HIV infection were seen but the HIV associated spongiform change and pallor of the white matter were present. HIV p24 antigen was detected in the brain by immunohistochemistry, the staining was described as diffuse. Moreover the patient had high HIV-1 proviral load in the brain (10^7 genome/µg DNA; Strappe et al., personal communication). Cell culture resulted in the isolation of adenovirus from the lymph node, spleen and lung and cytomegalovirus from the peripheral blood sample taken prior to death (Table 6.1). No opportunistic viruses were isolated from dorsal root ganglion or spinal cord.

Figure 6.3: Section of thoracic spinal cord showing vacuolar myelopathy in the posterior columns and the left dorso-lateral column. Haematoxylin and eosin stain; magnification X 4.

6.4 PHYLOGENETIC ANALYSIS

The maximum parsimony unrooted phylogenetic tree computed from approximately 19kb of LTR sequences derived from the peripheral blood, lymph node, spleen, lung, dorsal root ganglion and spinal cord is shown in Figure 6.4. The consensus tree had a characteristic polarised topography indicative of the presence of genetically distinct subgroups and possessed two major branches; one exclusively comprising the ganglion and spinal cord variants and the other comprising all the other variants. The polarisation between the nervous system variants and the remaining variants in the phylogenetic tree was observed in 100% of the bootstrap resamplings. The two clades comprising the non-nervous tissue variants were associated with low bootstrap values indicating that these variants exist in an equilibrium state. In order to assess the phylogenetic relationship between the variants detected in this patient with prototype HIV-1 LTR sequences, the LTRs from HIV-1 strains LAV (a T-cell tropic strain) and JR_{csf} (a CNS derived strain) have been included in these analyses. As expected both sequences are phylogenetically distinct from all the quasispecies obtained from the patient samples. This observation is not surprising since both strains have been isolated following in vitro culture. Nevertheless, the JR_{csf} LTR clustered more closely to the nervous tissue variants than to the variants present in other organs. These two prototype LTR variants were chosen because in transgenic mice experiments the JR_{csf} - LTR was able to drive heterologous expression in the nervous tissues whereas the LAV-LTR failed to direct expression in these tissues. The phylogenetic trees described above have also been constructed using Fitch-Margoliash and neighbor-joining methods and the polarised topology remained a constant feature (data not shown).

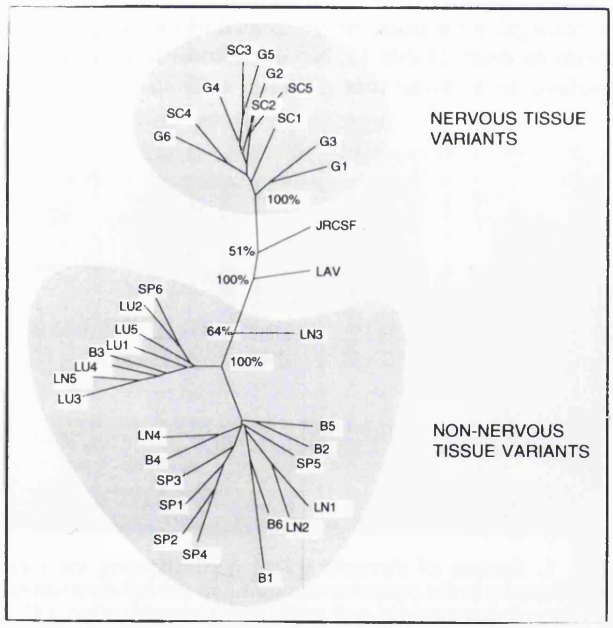


Figure 6.4: Unrooted maximum parsimony tree depicting phylogenetic relationships between the 33 LTR sequences obtained from the peripheral blood (variants B1-B6), lymph node (variants LN1-LN5), spleen (variants SP1-SP6), lung (variants LU1-LU5), spinal cord (variants SC1-SC5) and dorsal root ganglion (variants G1-G6) of the patient under investigation. This unrooted tree clearly shows that the nervous tissue variants cluster independently from the other tissue variants. The number of bootstrap resamplings yielding the branches is shown as a percentage. For comparison, the LAV and JR_{csf} prototype LTR sequences have been included.

6.5 MEDIAN DIVERGENCE AND MUTATION PATTERNS

The genetic heterogeneity within the different samples analysed has been determined by distance matrix methods (Table 6.1). The median heterogeneity in the tissues analysed ranged from 0.0094 to 0.017. There was no significant difference in intra-tissue median heterogeneity between the quasispecies present in any pairwise combination of tissues. These intraperiod diversities (one time point) are similar to that of patients 1, 2, and 3 described in Chapter 5 (at their latest time point): 0.0169, 0.0138, and 0.0246 respectively. These relatively low levels of divergence are in accordance with others (McNearney et al., 1995; Delassus et al., 1991).

To investigate the location of the distinct genetic features of the SC and G variants within the LTR sequence, the sequence alignment was analysed with respect to well defined functional elements (Figure 6.5). Several aspects of the multiple alignment should be noted.

Firstly, the sequence differences between the nervous system variants (SC and G) and the remaining variants were not randomly distributed, with divergence from the consensus sequence occurring more frequently within defined transcription factor binding sites (0.054/nucleotide versus 0.026/nucleotide; $p < 0.0001$). These sites included site B, the NF-AT binding site, USF, the up-regulatory element (URE) comprising the TCF-1 α binding site, Sp1, UBP/LBP binding sites and TAR.

TABLE 6.1: SUMMARY OF LTR DIVERSITY, HISTOLOGICAL FINDINGS AND VIRUS ISOLATION IN POST-MORTEM SAMPLES

ANATOMICAL SITE	HISTOLOGICAL FINDINGS	VIRUS ISOLATION	MEDIAN GENETIC DIVERSITY (range)
PBMC	Not applicable	Cytomegalovirus	0.0094 (0 - 0.024)
Lymph node	Generalised lymphoid depletion	Adenovirus	0.0132 (0 - 0.037)
Spleen	Generalised lymphoid depletion and congestion of red pulp	Adenovirus	0.0129 (0 - 0.027)
Lung	Bronchopneumonia and intra-alveolar haemorrhage. *P carinii* and yeast present.	Adenovirus	0.017 (0 - 0.04)
Dorsal root ganglion	Loss of ganglion cells	No virus isolated	0.0112 (0 - 0.017)
Spinal cord	Increased no. of *corpora amylacea*. Prominence of microglia, with microglial nodules in anterior horns at S1 level. Vacuolar myelopathy.	No virus isolated	0.015 (0 - 0.0226)

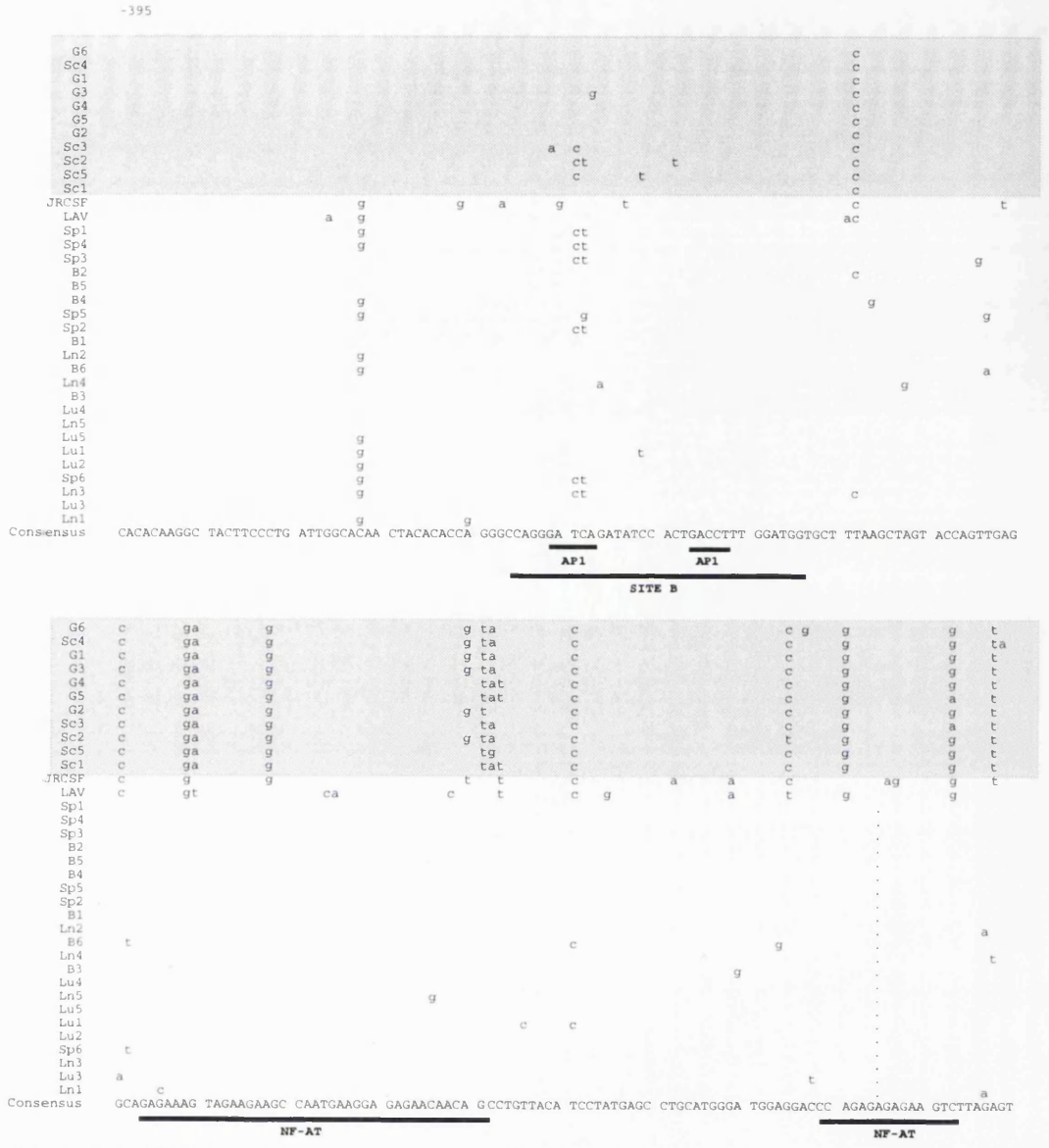


Figure 6.5: Alignment of the LTR sequences spanning positions -395 to +145 (relative to the start site of transcription) derived from the peripheral blood (BL), lymph node (LN), spleen (SP), lung (LU), spinal cord (SC) and dorsal root ganglion (G) of the patient under investigation. The viral and cellular transcription factor binding sites within the LTR are shown together with the junctions between the U3/R and R/U5 regions. This multiple alignment illustrates the extensive sequence divergence between the SC and G variants (highlighted in grey) and variants from the other tissues analysed. The nucleotide positions based on the prototype LAV sequence are shown for the beginning and end of the LTR fragment and the start of transcription. The LAV and JR_{csf} LTR are included for comparison at the interface of the nervous tissue and non-nervous tissue variants but do not contribute to the consensus sequence.

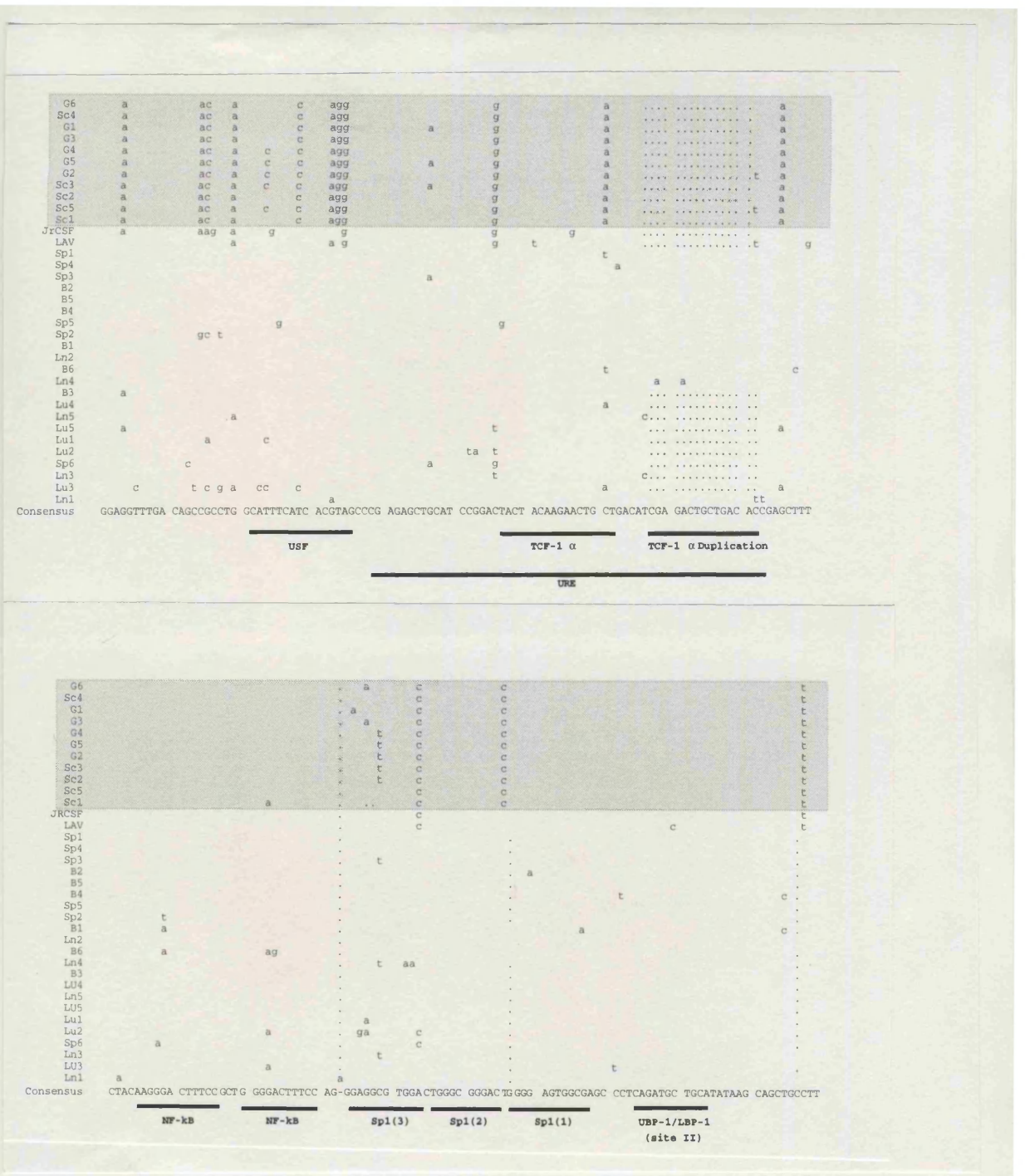


Figure 6.5 (cont.): Alignment of the LTR sequences spanning positions -395 to +145 (relative to the start site of transcription) derived from the peripheral blood (BL), lymph node (LN), spleen (SP), lung (LU), spinal cord (SC) and dorsal root ganlion (G) of the patient under investigation. The viral and cellular transcription factor binding sites within the LTR are shown together with the junctions between the U3/R and R/U5 regions. This multiple alignment illustrates the extensive sequence divergence between the SC and G variants (highlighted in grey) and variants from the other tissues analysed. The nucleotide positions based on the prototype LAV sequence are shown for the beginning and end of the LTR fragment and the start of transcription. The LAV and JR_{csf} LTR are included for comparison at the interface of the nervous tissue and non-nervous tissue variants but do not contribute to the consensus sequence.

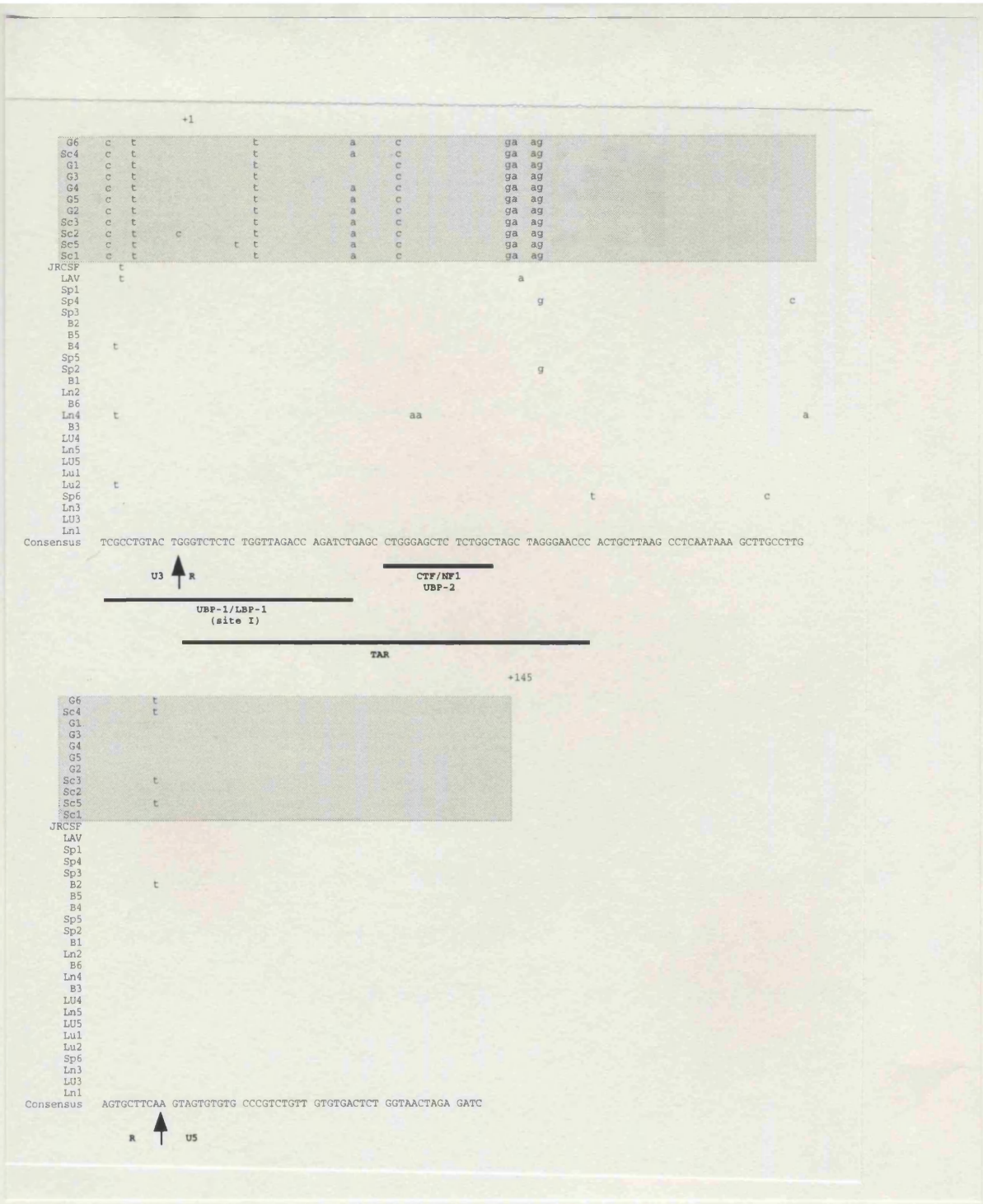


Figure 6.5 (cont.): Alignment of the LTR sequences spanning positions -395 to +145 (relative to the start site of transcription) derived from the peripheral blood (BL), lymph node (LN), spleen (SP), lung (LU), spinal cord (SC) and dorsal root ganlion (G) of the patient under investigation. The viral and cellular transcription factor binding sites within the LTR are shown together with the junctions between the U3/R and R/U5 regions. This multiple alignment illustrates the extensive sequence divergence between the SC and G variants (highlighted in grey) and variants from the other tissues analysed. The nucleotide positions based on the prototype LAV sequence are shown for the beginning and end of the LTR fragment and the start of transcription. The LAV and JR_{csf} LTR are included for comparison at the interface of the nervous tissue and non-nervous tissue variants but do not contribute to the consensus sequence.

Secondly, with the exception of variant SC1, all the SC and G NF-kB sites were identical to the prototypic NFκB sequences ("AGGGACTTTCC" and "GGGGACTTTCC") whereas a subset of the LN, LU, SP and B variants displayed 8 point mutations distributed between 4 polymorphic sites.

Thirdly, a 15 bp duplication was observed in 5/6 SP variants, 5/6 B variants and 2/5 LN variants that conferred a second potential binding site for the transcription factor TCF-1 alpha. This duplication was not observed in any of the lung or nervous tissue LTR variants.

Fourthly, the TAR sequences of the SC and G variants exhibited seven mutations, including a T to A change at position 25 of the TAR bulge and a T to C change at position 31 in the highly conserved TAR loop pentanucleotide CTGGG. These mutations do not affect the secondary structure as predicted using minimum free energy based computer programs (data not shown). However, the T25A mutation has been shown to reduce significantly *tat* mediated activation of transcription (Hamy et al., 1993); these TAR defective virions would be rescued via the TAR independent *tat* activation pathway.

6.6 DISCUSSION

I have sought to investigate the possible role of the HIV-1 LTR in the adaptation of HIV-1 strains in different tissues within the human host. By using a combination of DNA sequencing and phylogenetic analysis of HIV-1 LTR amplified by PCR from multiple tissues obtained at post-mortem examination of an AIDS patient I showed that a distinct HIV-1 LTR population was present in the spinal cord and dorsal root ganglion compared to that present in lung, spleen, lymph node and blood. Although the patient had no clinical evidence of neurological disease preceding death, histology of the brain and spinal cord revealed abnormalities characteristic of HIV-1 infection including extensive microglial nodules and vacuolar myelopathy. The nervous tissues and lymph node also contained high viral loads in comparison with other organs (unpublished data; Atkins et al., personal communication, manuscript in preparation). Analysis of the location of sequence variation between tissues revealed that deviation from the consensus sequence was more likely within transcription factor binding sites and was therefore consistent with the evolution of LTR variants most fit for survival in specific cell types. In comparison to the non-nervous tissue variants, the spinal cord and ganglion variants were more closely related to, but phylogenetically distinct from, a previously described LTR sequence (JR_{csf}) capable of driving HIV-1 expression in the CNS of transgenic mice and showed maintenance of prototypic NFκB binding sites but exhibited 7 mutations within the TAR sequence including T25A. On the basis of previous functional analyses, this mutation would be expected to significantly reduce *tat* activation (Hamy et al., 1993). These observations are consistent with the reported TAR independent activation of transcription mediated via NFκB (Taylor et al., 1994). Thus, the variants with mutated TAR sequences isolated from the G and SC of the

individual under study would be transcriptionally defective in the primary target cells of HIV-1, but not in microglial cells where *tat* transactivation can occur through the NFκB sites (Taylor et al., 1994).

The PCR method used in this study does not permit identification of the cell-type containing the HIV-1 LTR in the different tissues analysed. However, it is known that the major phenotype of infected cells in peripheral blood, lymphoid tissues and lung are CD4 bearing lymphocytes and macrophages. In addition, HIV infection of monocytes/macrophages, microglial cells and to a lesser extent astroglia and neurons have been reported in the central and peripheral nervous system (Watkins et al., 1990; Koenig et al., 1986; Vazeux et al., 1987; Price et al., 1988; Donaldson et al., 1994a). Thus, it is important to consider the underlying reasons that account for the presence of a unique population of LTR variants in the nervous tissues of the patient under investigation. It is unlikely that the data is confounded by specific LTR variants existing in linkage disequilibrium with certain env sequences (i.e. the data merely reflects the tissue tropism dictated by envelope sequences since the LTR quasispecies present in lung is phylogenetically distinct from that present in nervous tissues , although both have been shown to be macrophage dominated reservoirs for HIV. Furthermore, differences between envelope quasispecies both within the spleen and between spleen and blood have been reported from similar numbers of clones to those analysed in the present study (Delassus et al., 1992a), whereas the LTR in these two compartments cluster together in my data set (Figure 6.6). Finally, a longitudinal analysis of V3 region, nef region and U3 region of the LTR in four individuals has shown that unlike V3 which is characterised by increasing levels of sequence diversity from early time of infection to later stages of disease, LTR variation did not always increase in magnitude over time (McNearney et al., 1995).

Both these studies, confirm the extreme plasticity of HIV-1, but importantly have demonstrated that different loci of the genome can evolve independently and at different rates. This observation can be explained by the difference in selection pressures experienced by the different loci considered. I propose the following explanation for my findings in this Chapter. Low level infection of the CNS may have occurred early in the course of HIV infection followed by tissue specific adaptation via the LTR leading to HIV-1 variants optimally responsive to transcription factors predominating in the cells of the nervous tissues. This explanation is consistent with the increasing data highlighting the ability of HIV-1 LTR to sustain different (tissue-specific) pathways of transcriptional regulation. The data presented in this Chapter is also consistent with the results of a study of post-mortem samples showing that high HIV loads were predominantly present in lymphoid tissues of asymptomatic HIV-1 infected individuals whereas symptomatic individuals showed high viral burdens in both lymphoid and non-lymphoid tissues (Donaldson et al., 1994b). Hence, due to the nature of the seroconversion process in the patient analysed in the present study (concurrent symptomatic HCMV infection (Squire et al., 1992)), the immunosuppression may have been sufficiently pronounced at early stages of infection to enable HIV-1 to replicate in a large spectrum of organs and so play a direct role in pathogenesis over a relatively short time period (Squire et al., 1992; Pantaleo et al., 1993a; Frost and McLean, 1994). At present, the mechanism(s) by which alterations in the CNS of HIV-1 infected individuals occurs is still unclear but has been associated with expression of gp120 (Toggas et al., 1994) and the presence of high levels of unintegrated proviral DNA in the brains of patients with HIV encephalitis (Pang et al., 1990). Interestingly, the patient reported here also has unique codons related to zidovudine resistance in the brain as compared to other organs (P. Strappe; personal communication).

As a consequence of the dynamics of HIV-1 replication at all stages of disease, the virus exhibits extensive sequence variation leading to the rapid evolution of virus strains followed by the outgrowth of mutant viruses with selective advantages such as drug resistance (Kellam et al., 1994), escape from immune surveillance (Arendrup et al., 1993) and increased cytopathogenicity (de Jong et al., 1992b). The plasticity of the HIV-1 genome provide the virus with the ability to respond to selection pressures and adapt to the local environments (Coffin, 1995). The data presented in this chapter has revealed that a distinct group of HIV-1 LTR variants were present in the nervous system of the patient studied strengthening the hypothesis that the HIV-1 LTR may be an important determinant in tissue specific expression of HIV-1 following modulation in tropism dictated by envelope glycoprotein sequences.

CHAPTER 7

GENERAL DISCUSSION

7.1 OVERVIEW

The major aim of this thesis was to investigate the *in vivo* sequence variation of the HIV-1 LTR, both over time and in different anatomical organs. The study addressed 4 aspects of HIV-1 natural history and pathogenesis. First, to investigate the extent of heterogeneity of HIV-1 quasispecies at the LTR locus in infected patients. Second, to determine changes in the magnitude of sequence diversity from early times of infection to later stages of disease. Data on the V3 region of HIV-1 envelope was also obtained for 2 patients (Chapter 4) and compared to the findings on the LTR locus. Third, to investigate the relationship between quasispecies present in lymph node and that present in the peripheral blood using phylogenetic analyses and hence determine the evolution of LTR quasispecies over time. Fourth, to investigate the possible role of the LTR in tissue-specific adaptation of HIV-1 by sequence analysis of LTR quasispecies derived from multiple body sites.

The environment of continuous *de novo* virus infection, extensive replication and high CD4 cell turnover (Wei et al., 1995; Ho et al., 1995; Wain-Hobson, 1995a) provides an ideal setting for the generation of HIV variants. The multiple sources of generation of sequence diversity for HIV include the reverse transcriptase step with an *in vivo* nucleotide misincorporation in the order of 10^{-5} per base per cycle, and genetic recombination which occurs in up to 20% of replication events. Another possible source of generation of diversity is RNA pol II which has been shown to have a misincorporation rate of 10^{-5} per base per cycle *in vitro*. Many forces will affect the selection and persistence sequence variants: for example; the constraints on structure and function of proteins or nucleic acid (eg the TAR, the RRE regions of HIV-1) resulting in negative selection; the founder effect (the bottleneck transmission of a small number of sequences); the evolution of strains most

replication competent in the host; and the host's immune system.

The consequences of the high variability in the genetic composition of HIV are: a rapid intrapatient evolution through transmission bottlenecks, a rapid adaptation to new evolutionary niches, and the possibility of escaping the immune system. The wide range of HIV-1 variants which remain fully functional poses difficulties in the design of vaccines and of antiviral drugs to which the virus would remain sensitive.

The fluctuation of HIV variants in different body sites has been investigated by several groups. In 1991, Leigh-Brown, Simmonds and collaborators (Simmonds et al., 1991) showed that variants could first be detected as plasma virions, then subsequently isolated in the blood as proviral species. Similar data with more accurate temporal analysis were reported from the more recent study of Wei and collaborators who investigated the appearance rate of variants with resistant mutations to nevirapine in plasma and blood. At 14 days post-treatment, 100% of plasma virions carried the resistant mutation whereas the proviral DNA lags behind with between 50 -75% of the proviral DNA carrying the mutations after 180 days, representing evidence for a slow turnover of proviral sequences.

Since replication is required to generate diversity, it is logical to suppose that the site(s) of high level of replication is the site of generation of variation. Several studies have demonstrated that active replication occurs at all stages of HIV-1 infection in secondary lymphoid organs, and that viral load was higher in these organs than in peripheral blood (Pantaleo et al., 1991; Pantaleo et al., 1993a; Embretson et al., 1993; Tamalet et al., 1994; Lafeuillade et al., 1995; Pantaleo et al., 1995; Pantaleo et al., 1993b). I have investigated the role of lymphoid organs as

the site of evolution by using PCR and sequencing of multiple V3 region clones for 1 patient, and multiple LTR clones for 4 patients. The sequence data were analysed using software applications to generate sequence alignments, calculate median divergences of the quasispecies, and by the construction of phylogenetic trees to display the evolutionary relationship between variants. These methods were used to analyse the V3 region quasispecies from 2 patients (patients 0 and 1), to analyze the HIV-1 LTR natural variants from lymph nodes and peripheral blood from 4 individuals (patients 1 to 4), and the LTR quasispecies derived from post-mortem samples of lymph node, spleen, lung, dorsal root ganglion, and spinal cord as well as from peripheral blood of patient 4, who died of AIDS.

The patients analysed longitudinally comprised individuals who had a variety of lymph node histological architectures and CD4 cell numbers. Patient 2 had a lymph node with an intact histological structure at the time of biopsy, whereas patients 1, and 3 had follicular lysis and destruction of the histological structure of the lymph node. Several blood samples were obtained from these patients, including one at the time of biopsy, in addition to ones following biopsy and in the case of patient 1, 14 months prior to biopsy. The phylogenetic tree for the patient with an intact lymph node (patient 2) exhibited a polarised topology with two main branches, each almost exclusively consisting of lymph node or blood variants. This polarisation suggests that the lymph node variants can evolve independently from the blood variants. The structure of this tree also suggests that HIV-1 variants and HIV-1 infected cells may be sequestered in lymphoid organs. These results are consistent with the studies on viral load by PCR and *in situ* PCR (Pantaleo et al., 1991; Pantaleo et al., 1993a; Embretson et al., 1993; Donaldson et al., 1994b). For patient 1 who had a disrupted lymph node, however, a star-like tree was obtained

characteristic of closely related variants, with lymph node and blood variants clustering in common branches. The analysis of the V3 region quasispecies in lymph node and simultaneous blood of patient 1 was similar to that of the LTR locus, with lymph node and PBMC V3 variants present in common clades. The blood sample obtained 14 months prior to biopsy harboured LTR variants very closely related to each other whereas the blood variants at the time of biopsy and 6 months later are intermingled with lymph node variants in common branches. It is worth noting that several LTR variants from the blood follow-up sample were found in sub-branches clustering with a particular lymph node variant. This suggests that both lymph node and follow up blood variants had evolved from the same ancestor and in a similar fashion; or, more probably, that the blood variant evolved from the lymph node variant. Similar results were obtained for patient 3 with closely related variants. Patient 3 also had a disrupted lymph node histological structure.

Overall, these results suggest that patients with disrupted lymph node histology harboured genetically similar HIV-1 variants at the LTR and the V3 region loci, whereas, the patient with the intact lymph node structure harboured genetically distinct LTR variants in lymph node and peripheral blood, implying that an intact lymph node structure is necessary to sustain the compartmentalisation of HIV-1 infected cells. However, only one patient with an intact lymph node was available for analysis, and more data are required to confirm the hypothesis proposed above. However, in support of this hypothesis are the findings of a study on HIV-1 phenotypes present in lymph node and blood of 11 patients by Tamalet and colleagues (1994) (Tamalet et al., 1994). Patients with follicular hyperplasia but intact histological architecture of the lymph node, showed several phenotypic differences between the lymph node bulk variants and the blood variants. Firstly, the

infectious HIV-1 load in the lymph node of these patients was 1.74 log higher than that in the PBMCs, whereas patients with a disrupted lymph node architecture had a lymph node HIV-1 load only 0.49 log higher than that of the PBMCs. In addition, the ability of the bulk isolates to form syncytia in PBMC and MT2 cells revealed differences between lymph node and blood isolates only for patients with intact lymph node structure. Differences were also observed on the neutralising potency of autologous sera, with lymph node variants being resistant to neutralization. This study provides evidence for the role of the lymphoid organs as a site of replication and consequently a site of variation which is consistent with all the studies on PCR quantitation of viral load in lymphoid organs. Importantly, the findings of Tamalet and colleagues are in accordance with the LTR variation analysis presented in this thesis and discussed in the preceding section (see Chapters 5 and 6).

HIV tropism and virulence and their genetic determinants have been investigated by several groups (for review see Levy J.A 1994). Most studies have focused on the V3 region of HIV-1 envelope, and have shown the association of faster disease progression with the isolation of SI variants (see Section 1.3.1). The largest studies have been carried out by the group of Tersmette and Goudsmit (Boucher and Larder, 1994) who showed that during a 34-month follow-up, only 25% of infected individuals harbouring SI variants remained AIDS-free, whereas as many as 90% of the NSI carriers were AIDS-free. Despite the importance of these findings, several points should be noted. Firstly, SI carriers are a minority of HIV infected individuals; in the above study 22/166 (12%). Secondly, the SI and NSI phenotype is an *in vitro* characterisation. Thirdly, when the bulk isolate is characterised as SI, sequencing of multiple clones shows that only a minority carry the SI genetic determinants. Finally, the NSI variants can be extremely pathogenic in microglial cells

(macrophage lineage) in the brain, being able to form multinucleated giant cells *in vivo* (Donaldson et al., 1994a). As Goudsmit and colleagues have pointed out, the causal relationship between the presence of SI variants and disease progression is very difficult to demonstrate despite the clear association between SI variants and rapid progression. It is important to note that viral load has been shown to be the most important factor in disease progression and SI variants might be a mere reflection of increased viral load (Mellors et al., 1995).

The genetic determinant of the SI phenotype has been mapped to the V3 loop, with substitutions at positions 320 (D to Q)and 324 (D to N) to increase the overall charge associated with the V3 loop. A general methodology for the classification of SI or NSI based on the sequence of their V3 loop has been proposed by Simmonds and co-workers (Donaldson et al., 1994a). The method is based on the combination of the charge of the loop and the number of amino acid differences in the variant compared to the MN prototype V3 loop sequence. In chapter 4, I report the sequence analysis of the V3 region quasispecies derived from the peripheral blood of an individual undergoing HIV-1 seroconversion (patient 0). The results showed that the V3 region (aa249-376) was very homogeneous with only 5 variants present in the 15 analysed, with the V3 loop itself (aa296-330) displaying an even higher homogeneity with only 3 variants present. The V3 loop central motif of this individual was GPGK rather than the more common GPGR motif. On the basis of overall charge, all the V3 loop genotypes were predicted to be of the NSI phenotype. In contrast, patient 1 who had a CD4 count of 220 x 10^6/L, and CDC IVC disease, harboured an extremely heterogeneous V3 region (aa249-375) with no identical variants amongst the 23 clones analysed. Fifteen of the 23 variants were characterised by the deletion of amino acid 319 (relative to HXB2 prototype) and a

3 amino acid deletion from aa355 to aa357. In addition, two variants showed a 4 amino acid deletion (aa355-358). Patient 1 who harboured a V3 quasispecies characterised as NSI on the basis of charge distribution had a stable CD4 count of 200/µl over 4 months following the sample obtained for molecular analysis. The greater heterogeneity of the whole V3 region compared to that of the V3 loop demonstrates that the latter is under strong negative selection due to differential functional constraints relative to the flanking regions.

Moreover, the phylogenetic analysis of the V3 quasispecies from patient 1 revealed a tree with two main branches with bootstrap values of 97% (4 VLNs and 3 VBs) and 68% (3 VBs and 1 LN). The tree also comprised 12 other variants (6 VLNs and 6 VBs) closely related to each other forming a star phylogeny between the main two clades. This revealed two distinct V3 variants present in both lymph node and blood. The fact that overall, the 11 lymph node variants and the 12 blood variants were closely related clustering in common clades on the branches of the tree suggests the absence of independent evolution of HIV-1 in these two compartments (See Chapter 4, section 4.2.3). Alternatively, it is possible that the V3 region of the genome is independently evolving in the two anatomical sites but undergoing a convergent evolution. However, it is important to note that the samples from which the V3 region quasispecies were isolated, were also used to derive the LTR quasispecies. Interestingly, phylogenetic analysis of these LTR variants also showed the lymph node and blood variants to be closely related (see chapter 5), but unlike the V3 region, the tree's branches of the LTR variants were not associated with significant bootstrap value that would distinguish some variants from others. The above analysis represents evidence for the independent evolution of the V3 region and the LTR.

The generation of immune escape variants has also been proposed as a factor influencing the rate of disease progression (Nowak et al., 1995b). Analysis of 6 patients undergoing HIV-1 seroconversion demonstrated the emergence of oligoclonal expansion of T-cells with specific Vβ receptors, which were CD8+ve T-cells with HIV-1 specific toxicity (Pantaleo et al., 1994). Their data suggested that CTLs play an important role in the control of viral replication and virus dissemination during primary infection. Similar studies have been carried out by the groups of Koup and Ho. They showed that as soon as 1 to 3 weeks post-primary infection with HIV-1, CTLs can be detected in HIV infected individuals and that the increasing CTL frequencies are concomitant with a decrease in viral load (Koup et al., 1994). These data suggested that cellular immunity is involved in the initial control of virus replication in primary HIV-1 infection and indicates a role for CTL in protective immunity to HIV-1 *in vivo*. This early effective immune response is sustained during the asymptomatic stage of the infection, whilst a decrease in CTL frequencies is associated with late stage disease (Nowak et al., 1995b). The generation of escape variants has been proposed to account for this decrease in the cellular immune response and subsequent disease progression. There is some evidence for the appearance of HIV-1 escape variants during the period of clinical latency. A study investigating the CTL frequencies against a *gag* epitope restricted by HLA B27 in 2 patients as well as the *gag* proviral variants harboured by these individuals, showed that all the natural variants sequenced during the 14-months of follow-up were recognised by the CTLs of these patients and did not constitute escape variants (Meyerhans et al., 1991). However, another study has shown that 2 out of 8 individuals harboured PBMC variants that were able to escape CTL recognition through HLA B8 (Phillips et al., 1991). The extent to which these epitope variants constitute real CTL escape mutants *in vivo* is difficult to assess. Indeed an HIV-1

infected cell will present different epitopes through different HLA class I restrictions. For example, four epitopic regions in HIV-1 *nef* protein have been identified (aa73-82; aa83-97; aa113-128; and aa126-144) as multirestricted immunodominant regions recognised by CTLs(Culmann-Penciolelli et al., 1994). Each epitope was recognised under different HLA Class 1 restrictions. Thus, different HLA Class 1 molecules select distinct epitopes from HIV-1 proteins to stimulate CTL responses. Epitope regions conserved in sequence because of strong functional pressures would elicit substained CTL responses, but if amino acid changes in the CTL epitopes are tolerated by the virus, escape may arise. A mathematical model of the CTL response based on the above experimental findings and the theoretical understanding of the cellular arm of the immune system has been proposed (see Chapter 1, section 1.3.2 and Figure 1.10) (Nowak et al., 1995b)

HIV-1 is polytropic, and the plasticity of its genome allows the virus to adapt and successfully infect different cell types. Particular attention has been given to HIV-1 isolates from the nervous system (CNS variants) because of the early incidence of neurological disease in HIV-1 infected individuals. Cheng-Mayer and collaborators (1989) have shown that paired isolates from PBMC and cerebrospinal fluid from 6 individuals could be differentiated phenotypically: the CSF isolates were characterised by a better replication in primary macrophage culture whereas the blood isolates replicated better in T-cells (Cheng-Mayer et al., 1989). Genetic studies have focused on the V3 loop because it is thought to be the major immunodominant epitope for humoral immunity but more importantly because it has been shown to be a major determinant of HIV-1 tropism and syncitia-inducing phenotype. The V3 region of the envelope, as well as V4 and V5 regions have been shown to display a tissue-specific and host-determined evolution. CNS variants have

a common phenotype, they are macrophage/monocyte tropic and of the NSI phenotype, moreover, they are characterised by their ability to form multi-nucleated giant cells *in vivo* (Donaldson et al., 1994a). The primary determinant of macrophage/monocyte/microglia tropism for CNS variants is the V3 region of the envelope. However, once HIV infects a cell it could evolve at other sites to maximise replication. In this thesis, the sequence analysis of HIV-1 LTR variants in post-mortem tissues of one individual who died with AIDS was carried out in order to investigate the possible role of the LTR in tissue adaptation. Ideed, the LTR has been shown to be a determinant of tissue-specific expression for a number of retroviruses: eg Visna Virus, murine leukaemia virus, Rous sarcoma Virus, and HIV-1 (see Chapter 6, section 6.1). In the case of HIV-1, an NF-κB dependent but TAR independent *tat* transactivation has been demonstrated in microglial cells where TAR deleted variants were responsive to *tat* (Taylor et al., 1992; Bagasra et al., 1992; Taylor et al., 1994). The phylogenetic analysis described in Chapter 6, showed that the quasispecies present in multiple body sites formed a polarised tree with two distinct main branches characteristic of the presence of genetically distinct variants. The striking feature of this tree was the exclusive clustering of the spinal cord and ganglion variants in one clade whereas the non-nervous tissue variants ie: lymph node, blood, lung, and spleen natural variants clustered in the second clade. To investigate further genetic differences I analysed the multiple alignment of the LTR variants with respect to well known transcription factor binding sites. Several features of the alignment were noted: the base substitutions between the nervous and lymphoid variants were mainly in well characterised transcription binding sites, but more importantly, the TAR sequence of the LTR which is usually well conserved in most of the HIV-1 variants, had 7 mutations in the ganglion and spinal cord variants. Specifically, a T to A mutation at position 25 of TAR which would

significantly reduce *tat* activation (Hamy et al., 1993). Thus, these nervous tissue natural variants, which are impaired in their TAR mediated *tat* transactivation, would be rescued by the TAR independent *tat* transactivation mechanism described in microglial cells (Taylor et al., 1994). Consistent with this possibility is the high proviral load (10^5 genomes/μg DNA) in the post-mortem brain sample of this patient (Padraig Strappe et al., personal communication). Moreover, these nervous variants showed variability upstream of the NF-κB sites in the U3 region which has been shown to influence tissue specific expression (Corboy et al., 1992; Buzy et al., 1995). Thus, the results reported in chapter 6, suggest that differences in the nucleotide sequence of the LTRs of HIV-1 strains from different tissues may reflect the virus adapting to alter and optimise its tissue-specific gene expression in particular physiological environments. The results strengthen the hypothesis that the HIV-1 LTR may be an important determinant of tissue-specific expression of HIV-1 following modulation in tropism dictated by envelope glycoprotein sequences. However, the results reported in Chapter 6 are only from one patient and functional studies are required to confirm the interpretation of the results(see Section 7.2). Such investigations will provide further insight into HIV pathogenesis and neurovirulence.

In summary, high levels of HIV-1 replication and high turnover of CD4 cells occur at all stages of HIV-1 infection. The preferred sites of HIV-1 replication are the secondary lymphoid organs; indeed, lymphoid organs are a privileged site for efficient transactivation of the HIV-1 genome (upon immune activation of latently infected CD4+ cells), since high levels of stimulatory cytokines are present in this micro-environment. The results presented in this thesis, support the sequestration of HIV-1 infected cells in distinct anatomical sites and that the site of replication is

the site of variation, leading to independent evolution of quasispecies. However, the dynamics of the HIV-1 quasispecies are extremely complex and difficult to predict because the *in vivo* selection forces driving evolution vary during the course of infection. For instance, the immune system elicits responses that are discontinuous (different dominant epitopes) during the disease course, and the replication rate of the virus is also variable and influenced by an array of factors which may differ from individual to individual.

The extensive studies of HIV-1 sequence variation and phenotype have demonstrated that no single region of the HIV-1 genome dictates the phenotypic diversity of the virus, ie the *in vivo* effects of altered tropism, gene expression, and antigenic variation on disease progression. A high level of viral replication appears to be a fundamental factor in AIDS pathogenesis and as a consequence it is difficult to state that a particular change in phenotype is of crucial importance for pathogenesis because the multiple factors influencing disease progression are all interrelated.

7.2 FUTURE WORK

The work in this thesis has described the co-existence of genetically distinct HIV-1 LTR variants in multiple tissues of HIV-1 infected individuals. In particular, LTR variants in the nervous tissues of an AIDS patient were divergent from those from lymphoid tissues, suggesting that the LTR may have a role in tissue adaptation, with specific nucleotide changes occuring in transcription factor binding sites resulting in optimisation of HIV-1 gene expression in different tissues. Several other studies have reported similar findings on the third variable region of the envelope gene of HIV-1, demonstrating that the V3 region genotypes from nervous tissues are of the

NSI and monocytotropic phenotypes, which suggests specific virus entry in nervous tissue cells, and confirming immunocytochemical and PCR results showing that the predominant HIV infected cells in these tissues are microglia, a nervous tissue-specific cell of the monocytic lineage. Several studies including those described in this thesis have shown that different regions of the HIV-1 genome evolve independently, but to formally prove that LTR and *env* sequences (specifically V3) do not exist in linkage disequilibrium, long PCR could be performed encompassing the *env* gene through to the U5 region of the 3' LTR. Sequence analysis of multiple clones of these long amplicons (*env* and LTR from the same clone) will enable the extent of linkage between the two loci to be evaluated. The drawback of long PCR for sequence analysis is the increased likeliness of recombination events, an alternative to long PCR would be the amplification of single molecules after end-point dilution.

Several experiments could be performed to analyse the cell-specific transcription potencies of the LTR variants described in this thesis. LTR-directed expression of reporter genes such as β-galactosidase, luciferase, CAT following transfection of different cell types, especially lymphoid versus nervous, should be carried out to confirm the hypothesis that LTR variants are most suited to expression in specific cell types. Moreover, the use of the LTRs isolated *ex vivo* in transgenic mice could be carried out to investigate the specific pattern of expression in different tissues. Indeed, the requirement for an interaction between the virus and its cellular receptor is bypassed in the transgenic mice models, therefore allowing the direct investigation of the tissue-specific transcriptional regulation and viral gene expression of LTR variants.

Differential expression *via* the LTR suggests that nuclear proteins acting as transcription factors may be unique in nervous tissues. Experiments could be performed to identify and characterise these factors, mainly through cellular protein binding assays (DNA mobility shift assays) using probes derived from the different LTR natural variants encompassing the known transcripton binding site and nuclear extracts from different cell lineages. It would also be of interest to identify the sequences in the LTRs that may be involved in nervous tissue-specific transcription, this could be achieved by DNAase I footprinting assays.

Finally, similar functional studies to those described above should be carried out in different cell lineages known to be infected by HIV-1 *in vivo*, such as, epithelial and endothelial cells from different organs, dendritic cells, Langerhans cells, and Kupffer cells. Since transcription factors act as a regulatory network, the aforementioned investigations would provide insight into the different pathways of transcriptional regulation of HIV-1 expression in multiple tissues and provide a functional complement to the data described in this thesis.

REFERENCES

Ahmad, N., Baroudy, B.M., Baker, R.C., and Chappey, C. (1995). Genetic analysis of human immunodeficiency virus type 1 envelope V3 region isolates from mothers and infants after perinatal transmission. Journal of Virology *69*, 1001-1012.

Ait-Khaled M, McLaughlin J.E, Johnson M.A, and Emery V.C (1995). Dinstinct HIV-1 long terminal repeat quasispecies present in nervous tissues compared to that in lung, blood and lymphoid tissues of an AIDS patient. AIDS *9*, 675-683.

Ait-Khaled, M. and Emery, V.C. (1993). Homogeneity of human immunodeficiency virus V3 loop sequences at seroconversion. Journal of Medical Virology *41*, 270-274.

Ait-Khaled, M. and Emery, V.C. (1994). Phylogenetic relationship between human immunodeficiency virus type 1 (HIV-1) long terminal repeat natural variants present in the lymph node and peripheral blood of three HIV-1-infected individuals. Journal of General Virology *75*, 1615-1621.

Alcami, J., Lain de Lera, T., Folgueira, L., Pedroza, M.A., Jacque, J.M., Bachelerie, F., Noriega, A.R., Hay, R.T., Harrich, D., Gaynor, R.B., and et al. (1995). Absolute dependence on kappa B responsive elements for initiation and Tat-mediated amplification of HIV transcription in blood CD4 T lymphocytes. EMBO Journal *14*, 1552-1560.

Alizon, M. and Dragic, T. (1994). CD26 antigen and HIV fusion? [letter; comment]. Science *264*, 1161-2; discussion 1162-5.

Altman D.G (1993). Comparing groups-continuous data. In Practical statistics for medical research. Altman D.G, ed. (chapman & hall), pp. 179-228.

Anand, R., Reed, C., Forlenza, S., Siegal, F., Cheung, T., and Moore, J. (1987). Non-cytocidal natural variants of Human Immunodeficiency Virus Isolated from AIDS patients with neurological disorders. Lancet 234-238.

Arendrup, M., Sonnerborg, A., Svennerholm, B., Akerblom, L., Nielsen, C., Clausen, H., Olofsson, S., Nielsen, J.O., and Hansen, J.E.S. (1993). Neutralizing antibody response during human immunodeficiency virus type 1 infection: type and group specificity and viral escape. Journal of General Virology *74*, 855-863.

Ascher, M.S., Sheppard, H.W., Anderson, R.W., Krowka, J.F., and Bremermann, H.J. (1995). HIV results in the frame. Paradox remains [letter; comment]. Nature *375*, 196; discussion 198

Ashorn, P.A., Berger, E.A., and Moss, B. (1990). Human immunodeficiency virus envelope glycoprotein/CD4-mediated fusion of nonprimate cells with human cells. Journal of Virology *64*, 2149-2156.

Asjo, B., Morfeldt-Manson, L., Albert, J., Biberfeld, G., Karlsson, A., Lidman, K, and Fenyo, E.M. (1986). Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection. Lancet *2*, 660-662.

Bagasra, O., Khalili, K., Seshamma, T., Taylor, J.P., and Pomerantz, R.J. (1992). TAR-Independent Replication of Human Immunodeficiency Virus Type 1 in Glial Cells. Journal of Virology *66*, 7522-7528.

Balfe, P., Simmonds, P., Ludlam, C.A., Bishop, J.O., and Brown, A.J. (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 mutations. Journal of Virology *64*, 6221-6233.

Ball, J.K., Holmes, E.C., Whitwell, H., and Desselberger, U. (1994). Genomic variation of human immunodeficiency virus type 1 (HIV-1): molecular analyses of HIV-1 in sequential blood samples and various organs obtained at autopsy. Journal of General Virology *75*, 867-879.

Barr, P.J. (1991). Mammalian subtilisins: the long-sought dibasic processing endoproteases. Cell *66*, 1-3.

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

Barry, P.A., Pratt-Lowe, E., Unger, R.E., and Luciw, P.A. (1991). Cellular factors regulate transactivation of human immunodeficiency virus type 1. Journal of Virology *65*, 1392-1399.

Bedinger, P., Moriarty, A., von Borstel, R.C., 2d, Donovan, N.J., Steimer, K.S., and Littman, D.R. (1988). Internalization of the human immunodeficiency virus does not require the cytoplasmic domain of CD4. Nature *334*, 162-165.

Bell, J. (1989). The polymerase chain reaction. Immunology Today *10*, 351-355.

Ben Artzi, H. and Panet, A. (1996). Molecular analysis of the reverse transcription process of the HIV genome. International Antiviral News *4*, 11-13.

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

Berkhout, B., Gatignol, A., Rabson, A.B., and Jeang, K.T. (1990). TAR-independent activation of the HIV-1 LTR: evidence that tat requires specific regions of the promoter. Cell *62*, 757-767.

Berkhout, B. and Jeang, K.T. (1992). Functional roles for the TATA promoter and enhancers in basal and Tat-induced expression of the human immunodeficiency virus type 1 long terminal repeat. Journal of Virology *66*, 139-149.

Berkowitz, R.D. and Goff, S.P. (1994). Analysis of binding elements in the human immunodeficiency virus type 1 genomic RNA and nucleocapsid protein. Virology *202*, 233-246.

Boucher C.A.B and Larder B.A (1994). Viral Variation and Therapeutic Strategies in HIV Infection. Bulletin, MediTech Media Ltd, Cambridge, UK.

Briant, L., Wade, C.M., Puel, J., Brown, A.J., and Guyader, M. (1995). Analysis of envelope sequence variants suggests multiple mechanisms of mother-to-child transmission of human immunodeficiency virus type 1. Journal of Virology *69*, 3778-3788.

Broder, C.C., Nussbaum, O., Gutheil, W.G., Bachovchin, W.W., and Berger, E.A. (1994). CD26 antigen and HIV fusion?. Science *264*, 1156-9; discussion 1162-5.

Brown, A.J. (1991). Sequence variability in human immunodeficiency viruses: pattern and process in viral evolution. AIDS *5 Suppl 2*, S35-42.

Brown, A.L. and Monaghan, P. (1988). Evolution of the structural proteins of human immunodeficiency virus: selective constraints on nucleotide substitution. AIDS Research & Human Retroviruses *4*, 399-407.

Buianouckas, F.R. (1995). HIV results in the frame. HIV an illusion. Nature *375*, 197.

Bukrinsky, M., Manogue, K., and Cerami, A. (1995). HIV results in the frame. Other approaches. Nature *375*, 195-6.

Buzy, J.M., Lindstrom, L.M., Zink, M.C., and Clements, J.E. (1995). HIV-1 in the Developing CNS: Developmental Differences in Gene Expression. Virology *210*, 361-371.

Callebaut, C., Krust, B., Jacotot, E., and Hovanessian, A.G. (1993). T cell activation antigen, CD26, as a cofactor for entry of HIV in CD4+ cells. Science *262*, 2045-2050.

Camerini, D., Planelles, V., and Chen, I.S. (1994). CD26 antigen and HIV fusion? Science *264*, 1160-1.

Cariello, N.F., Swenberg, J.A., and Skopek, T.R. (1991). Fidelity of Thermococcus litoralis DNA polymerase (Vent) in PCR determined by denaturing gradient gel electrophoresis. Nucleic Acids Research *19*, 4193-4198.

Chakrabarti, L., Cumont, M.C., Montagnier, L., and Hurtrel, B. (1994a). Variable Course of Primary Simian Immunodeficiency Virus Infection in Lymph Nodes: Relation to Disease Progression. Journal of Virology *68*, 6634-6642.

Chakrabarti, L., Cumont, M.C., Montagnier, L., and Hurtrel, B. (1994b). Kinetics of primary SIV infection in lymph nodes. Journal of Medical Primatology *23*, 117-124.

Chakrabarti, L., Isola, P., Cumont, M.C., Claessens-Maire, M.A., Hurtrel, M., Montagnier, L., and Hurtrel, B. (1994c). Early stages of simian immunodeficiency virus infection in lymph nodes. Evidence for high viral load and successive populations of target cells. American Journal of Pathology *144*, 1226-1237.

Chandrasekhar, K., Profy, A.T., and Dyson, H.J. (1991). Solution conformational preferences of immunogenic peptides derived from the principal neutralizing determinant of the HIV-1 envelope glycoprotein gp120. Biochemistry *30*, 9187-9194.

Charneau, P., Borman, A.M., Quillent, C., Guetard, D., Chamaret, S., Cohen, J., Remy, G., Montagnier, L., and Clavel, F. (1994). Isolation and envelope sequence of a highly divergent HIV-1 isolate: definition of a new HIV-1 group. Virology *205*, 247-253.

Chattopadhyay, S.K., Cloyd, M.W., Linemeyer, D.L., Lander, M.R., Rands, E., Lowy, and DR. (1982). Cellular origin and role of mink cell focus-forming viruses in murine thymic lymphomas. Nature *295*, 25-31.

Cheng-Mayer, C., Seto, D., Tateno, M., and Levy, J.A. (1988). Biologic features of HIV-1 that correlate with virulence in the host. Science *240*, 80-82.

Cheng-Mayer, C., Weiss, C., Seto, D., and Levy, J.A. (1989). Isolates of human immunodeficiency virus type 1 from the brain may constitute a special group of the AIDS virus. Proceedings of the National Academy of Sciences, USA *86*, 8575-8579.

Cheng-Mayer, C., Quiroga, M., Tung, J.W., Dina, D., and Levy, J.A. (1990). Viral determinants of human immunodeficiency virus type 1 T-cell or macrophage tropism, cytopathogenicity, and CD4 antigen modulation. Journal of Virology *64*, 4390-4398.

Chesebro, B., Nishio, J., Perryman, S., Cann, A., O'Brien, W., Chen, I.S., Wehrly, and K. (1991). Identification of human immunodeficiency virus envelope gene sequences influencing viral entry into CD4-positive HeLa cells, T-leukemia cells, and macrophages. Journal of Virology *65*, 5782-5789.

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

Cheynier, R., Meyerhans, A., Asjo, B., and Wain-Hobson, S. (1990). Temporal fluctuations in HIV quasispecies in vivo are not reflected by sequential HIV isolations. Developments in Biological Standardization *72*, 349-353.

Cheynier, R., Henrichwark, S., Hadida, F., Pelletier, E., Oksenhendler, E., Autran, B., and Wain-Hobson, S. (1994). HIV and T cell expansion in splenic white pulps is accompanied by infiltration of HIV-specific cytotoxic T lymphocytes. Cell *78*, 373-387.

Chiang, C.M., Ge, H., Wang, Z., Hoffmann, A., and Roeder, R.G. (1993). Unique TATA-binding protein-containing complexes and cofactors involved in transcription by RNA polymerases II and III. EMBO Journal *12*, 2749-2762.

Cichutek, K., Norley, S., Linde, R., Kreuz, W., Gahr, M., Lower, J., von, Wangenheim, G., and Kurth, R. (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.

Cichutek, K., Merget, H., Norley, S., Linde, R., Kreuz, W., Gahr, M., and Kurth, R. (1992a). Development of a quasispecies of human immunodeficiency virus type 1 in vivo. Proceedings of the National Academy of Sciences, USA *89*, 7365-7369.

Cichutek, K., Merget, H., Norley, S., Linde, R., Kreuz, W., Gahr, M., and Kurth, R. (1992b). Development of a quasispecies of human immunodeficiency virus type 1 in vivo. Proceedings of the National Academy of Sciences USA *89*, 7365-7369.

Cimino, G.D., Metchette, K., Isaacs, S.T., and Zhu, Y.S. (1990). More false-positive problems [letter; comment]. Nature *345*, 773-774.

Clark, S.J., Saag, M.S., Decker, W.D., Campbell-Hill, S., Roberson, J.L., VeldkampPJ., Kappes, J.C., Hahn, B.H., and Shaw, G.M. (1991). High titers of cytopathic virus in plasma of patients with symptomatic primary HIV-1 infection [see comments]. New England Journal of Medicine *324*, 954-960.

Clements, G.J., Price-Jones, M.J., Stephens, P.E., Sutton, C., Schulz, T.F., Clapham, P.R., McKeating, J.A., McClure, M.O., Thomson, S., Marsh, M., 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 Research & Human Retroviruses *7*, 3-16.

Clewley, J.P. (1989). The polymerase chain reaction, a review of the practical limitations for human immunodeficiency virus diagnosis. Journal of Virological Methods *25*, 179-188.

Coffin, J.M. (1990). Retroviridae and Their Replication. In Fields VIROLOGY. B.N. Fields and D.M. Knipe, eds. (Raven press), pp. 1437-1500.

Coffin, J.M. (1995). HIV Population Dynamics in Vivo: Implications for Genetic Variation, Pathogenesis, and Therapy. Science *267*, 483-489.

Collman, R., Balliet, J.W., Gregory, S.A., Friedman, H., Kolson, D.L., Nathanson, N, and Srinivasan, A. (1992). An infectious molecular clone of an unusual macrophage-tropic and highly cytopathic strain of human immunodeficiency virus type. Journal of Virology *66*, 7517-7521.

Cooney, A.J., Tsai, S.Y., O'Malley, B.W., and Tsai, M.J. (1991). Chicken ovalbumin upstream promoter transcription factor binds to a negative regulatory region in the human immunodeficiency virus type 1 long terminal repeat. Journal of Virology *65*, 2853-2860.

Cooper, D.A., Gold, J., Maclean, P., Donovan, B., Finlayson, R., Barnes, T.G., Michelmore, H.M., Brooke, P., and Penny, R. (1985). Acute AIDS retrovirus infection. Definition of a clinical illness associated with seroconversion. Lancet *1*, 537-540.

Corboy, J.R., Buzy, J.M., Zink, M.C., and Clements, J.E. (1992). Expression directed from HIV long terminal repeats in the central nervous system of transgenic mice. Science *258*, 1804-1808.

Cullen, B.R. and Greene, W.C. (1989). Regulatory pathways governing HIV-1 replication. Cell *58*, 423-426.

Culmann-Penciolelli, B., Lamhamedi-Cherradi, S., Couillin, I., Guegan, N., Levy, J.P., Guillet, J.G., and Gomard, E. (1994). Identification of Multirestricted Immunodominant Regions Recognised by Cytolytic T Lymphocytes in the Human Immunodeficiency Virus Type 1 Nef Protein. Journal of Virology *68*, 7336-7343.

Daar, E.S., Moudgil, T., Meyer, R.D., and Ho, D.D. (1991). Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection [see comments]. New England Journal of Medicine *324*, 961-964.

Dalgleish A.G. (1996). The CD4 Loss in AIDS Patients is Immunopathologically Mediated. Reviews in Medical Virology (in press).

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

de Jong, J.J., de Ronde, A., Keulen, W., Tersmette, M., and Goudsmit, J. (1992a). Minimal Requirements for the Human Immunodeficiency Virus Type 1 V3 Domain To Support the Syncytium-Inducing Phenotype: Analysis by Single Amino Acid Substitution. Journal of Virology *66*, 6777-6780.

de Jong, J.J., Goudsmit, J., Keulen, W., Klaver, B., Krone, W., Tersmette, M., and de Ronde, A. (1992b). Human Immunodeficiency Virus Type 1 Clones Chimeric for the Envelope V3 Domain Differ in Syncytium Formation and Replication Capacity. Journal of Virology *66(2)*, 757-765.

Delassus, S., Cheynier, R., and Wain-Hobson, S. (1991). Evolution of human immunodeficiency virus type 1 nef and long terminal repeat sequences over 4 years in vivo and in vitro. Journal of Virology *65*, 225-231.

Delassus, S., Cheynier, R., and Wain-Hobson, S. (1992a). Nonhomogeneous Distribution of Human Immunodeficiency Virus Type 1 Proviruses in the Spleen. Journal of Virology *66*, 5642-5645.

Delassus, S., Meyerhans, A., Cheynier, R., and Wain-Hobson, S. (1992b). Absence of selection of HIV-1 variants in vivo based on transcription/transactivation during progression to AIDS. Virology *188*, 811-818.

Delling, U., Reid, L.S., Barnett, R.W., Ma, M.Y., Climie, S., Sumner-Smith, M., and Sonenberg, N. (1992). Conserved nucleotides in the TAR RNA stem of human immunodeficiency virus type 1 are critical for Tat binding and trans activation: model for TAR RNA tertiary structure. Journal of Virology *66*, 3018-3025.

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

Di Fagagna, F.D., Marzio, G., Gutierrez, M.I., Kang, L.Y., Falaschi, A., and Giacca, M. (1995). Molecular and functional interactions of transcription factor USF with the long terminal repeat of human immunodeficiency virus type 1. Journal of Virology *69*, 2765-2775.

Dimitrov, D.S. and Martin, M.A. (1995). HIV results in the frame. CD4+ cell turnover. Nature *375*, 194-5.

Dingwall, C., Ernberg, I., Gait, M.J., Green, S.M., Heaphy, S., Karn, J., Lowe, A.D., Singh, M., and Skinner, M.A. (1990). HIV-1 tat protein stimulates transcription by binding to a U-rich bulge in the stem of the TAR RNA structure. EMBO Journal *9*, 4145-4153.

Donaldson, Y.K., Bell, J.E., Holmes, E.C., Hughes, E.S., Brown, H.K., and Simmonds, P. (1994a). In Vivo Distribution And Cytopathology of Variants of HIV-1 Showing Restricted Sequence Variability in the V3 loop. Journal of Virology *68(9)*, 5991-6005.

Donaldson, Y.K., Bell, J.E., Ironside, J.W., Brettle, R.P., Robertson, J.R., and Simmonds, P. (1994b). Redistribution of HIV outside the lymphoid system with onset of AIDS. Lancet *343*, 383-385.

Dragic, T., Charneau, P., Clavel, F., and Alizon, M. (1992). Complementation of murine cells for human immunodeficiency virus envelope/CD4-mediated fusion in human/murine heterokaryons. Journal of Virology *66*, 4794-4802.

Duesberg, P. and Bialy, H. (1995). HIV results in the frame. HIV an illusion . Nature *375*, 197.

Eiden, L.E. and Lifson, J.D. (1992). HIV interactions with CD4: a continuum of conformations and consequences. Immunology Today *13*, 201-206.

Embretson, J., Zupancic, M., Ribas, J.L., Burke, A., Racz, P., Tenner-Racz, K., and Haase, A.T. (1993). Massive covert infection of helper T-lymphocytes and macrophages by HIV during the incubation period of AIDS. Nature *362*, 359-362.

Engelman, A., Mizuuchi, K., and Craigie, R. (1991). HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. Cell *67*, 1211-1221.

Englund, G., Hoggan, M.D., Theodore, T.S., and Martin, M.A. (1991). A novel HIV-1 isolate containing alterations affecting the NF-kappa B element. Virology *181*, 150-157.

Ensoli, F., Ensoli, B., and Thiele, C.J. (1994). HIV-1 gene expression and replication in neuronal and glial cell lines with immature phenotype: effects of nerve growth factor. Virology *200*, 668-676.

Epstein, L.G., Kuiken, C., Blumberg, B.M., Hartman, S., Sharer, L.R., and Clement, M. (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.

Felsenstein, J. (1988). Phylogenies from molecular sequences: inference and reliability. Annual Review of Genetics *22*, 521-565.

Fenyo, E.M., Albert, J., and Asjo, B. (1989). Replicative capacity, cytopathic effect and cell tropism of HIV. AIDS *3 Suppl 1*, S5-12.

Folks, T.M., Clouse, K.A., Justement, J., Rabson, A., Duh, E., Kehrl, J.H., Fauci, and AS. (1989). Tumor necrosis factor alpha induces expression of human immunodeficiency virus in a chronically infected T-cell clone. Proceedings of the National Academy of Sciences, USA *86*, 2365-2368.

Fouchier, R.A.M., Groenink, M., Kootstra, N.A., Tersmette, M., Huisman, H.G., Miedema, F., and Schuitemaker, H. (1992). Phenotype-Associated Sequence Variation in the Third Variable Domain of the Human Immunodeficiency Virus Type 1 gp120 Molecule. Journal of Virology *66*, 3183-3187.

Fox, J.C., Ait-Khaled, M., Webster, A., and Emery, V.C. (1991). Eliminating PCR contamination: is UV irradiation the answer? Journal of Virological Methods *33*, 375-382.

Frost, S.D.W. and McLean, A.R. (1994). Germinal Centre Destruction as a Major Pathway of HIV Pathogenesis. Journal of Acquired Immune Deficiency Syndromes *7*, 236-244.

Fu, X.D. (1993). Specific commitment of different pre-mRNAs to splicing by single SR proteins. Nature *365*, 82-85.

Gao, F., Yue, L., White, A.T., Pappas, P.G., Barchue, J., Hanson, A.P., Greene, B.M., Sharp, P.M., Shaw, G.M., and Hahn, B.H. (1992). Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. Nature *358*, 495-499.

Gaynor, R. (1992). Cellular transcription factors involved in the regulation of HIV-1 gene expression. AIDS *6*, 347-363.

Ghazal, P., Young, J., Giulietti, E., DeMattei, C., Garcia, J., Gaynor, R., Stenberg, R.M., and Nelson, J.A. (1991). A discrete cis element in the human immunodeficiency virus long terminal repeat mediates synergistic trans activation by cytomegalovirus immediate-early proteins. Journal of Virology *65*, 6735-6742.

Golub, E.I., Li, G., and Volsky, D.J. (1990). Differences in the Basal Activity of the Long Terminal Repeat Determine Different Replicative Capacities of Two Closely Related Human Immunodeficiency Virus Type 1 Isolates. Journal of Virology *64(8)*, 3654-3660.

Goodenow, M., Huet, T., Saurin, W., Kwok, S., Sninsky, J., and Wain-Hobson, S. (1989). HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. Journal of Acquired Immune Deficiency Syndromes *2*, 344-352.

Griffin, G.E., Leung, K., Folks, T.M., Kunkel, S., and Nabel, G.J. (1989). Activation of HIV gene expression during monocyte differentiation by induction of NF-kappa B. Nature *339*, 70-73.

Grimaila, R.J., Fuller, B.A., Rennert, P.D., Nelson, M.B., Hammarskjold, M.L., Potts, B, Murray, M., Putney, S.D., and Gray, G. (1992). Mutations in the principal neutralization determinant of human immunodeficiency virus type 1 affect syncytium formation, virus infectivity, growth kinetics, and neutralization. Journal of Virology *66*, 1875-1883.

Gurtler, L.G., Hauser, P.H., Eberle, J., von Brunn, A., Knapp, S., Zekeng, L., Tsague, J.M., and Kaptue, L. (1994). A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. Journal of Virology *68*, 1581-1585.

Hamy, F., Asseline, U., Grasby, J., Iwai, S., Pritchard, C., Slim, G., Butler, PJ, Karn, J., and Gait, M.J. (1993). Hydrogen-bonding contacts in the major groove are required for human immunodeficiency virus type-1 tat protein recognition of TAR RNA. Journal of Molecular Biology *230*, 111-123.

Harrich, D., Garcia, J., Mitsuyasu, R., and Gaynor, R. (1990). TAR independent activation of the human immunodeficiency virus in phorbol ester stimulated T lymphocytes. EMBO Journal *9*, 4417-4423.

Harrich, D., Mavankal, G., Mette-Snider, A., and Gaynor, R.B. (1995). Human immunodeficiency virus type 1 TAR element revertant viruses define RNA structures required for efficient viral gene expression and replication. Journal of Virology *69*, 4906-4913.

Hart, C.E., Ou, C.Y., Galphin, J.C., Moore, J., Bacheler, L.T., Wasmuth, J.J., Petteway, S.R., Jr., and Schochetman, G. (1989). Human chromosome 12 is required for elevated HIV-1 expression in human-hamster hybrid cells. Science *246*, 488-491.

Hart, C.E., Galphin, J.C., Westhafer, M.A., and Schochetman, G. (1993). TAR loop-dependent human immunodeficiency virus trans activation requires factors encoded on human chromosome 12. Journal of Virology *67*, 5020-5024.

Hart, C.E., Saltarelli, M.J., Galphin, J.C., and Schochetman, G. (1995). A human chromosome 12-associated 83-kilodalton cellular protein specifically binds to the loop region of human immunodeficiency virus Type 1 *trans*-activation response element RNA. Journal of Virology *69*, 6593-6599.

Haynes B.F., Pantaleo G., and Fauci A.S. (1996). Towards an Understanding of the Correlates of Protective Immunity to HIV infection. Science *271*, 324-328.

Heath, S.L., Grant Tew, j., Tew, j.g., szakal, a.k., and burton, g.f. (1995). Follicular dendritic cells and human immunodeficiency virus infectivity. Nature *337*, 740-744.

Herrmann, C.H. and Rice, A.P. (1993). Specific interaction of the human immunodeficiency virus Tat proteins with a cellular protein kinase. Virology *197*, 601-608.

Herrmann, C.H. and Rice, A.P. (1995). Lentivirus Tat Proteins Specifically Associate with a Cellular Protein Kinase, TAK, That Hyperphosphorylates the Carboxyl-Terminal Domain of the Large Subunit of RNA Polymerase II: Candidate for a Tat Cofactor. Journal of Virology *69(3)*, 1612-1620.

Hillis, D.M., Huelsenbeck, J.P., and Swofford, D.L. (1994). Hobgoblin of phylogenetics?. Nature *369*, 363-364.

Ho, D.D., Neumann, A.U., Perelson, A.S., Chen, W., Leonard, J.M., and Markowitz, M. (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature *373*, 123-126.

Holland, C.A., Wozney, J., Chatis, P.A., Hopkins, N., and Hartley, J.W. (1985). Construction of recombinants between molecular clones of murine retrovirus MCF 247 and Akv: determinant of an in vitro host range property that maps in the long terminal repeat. Journal of Virology *53*, 152-157.

Holmes, E.C., Zhang, L.Q., Simmonds, P., Ludlam, C.A., and Brown, A.J.L. (1992). Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. Proceedings of the National Academy of Sciences, USA *89*, 4835-4839.

Holmes, E.C., Zhang, L.Q., Robertson, P., Cleland, A., Harvey, E., Simmonds, P., and Brown, A.J.L. (1995). The Molecular Epidemiology of Human Immunodeficiency Virus Type 1 in Edinburgh. Journal of Infectious Diseases *171*, 45-53.

Hu, W.S. and Temin, H.M. (1990). Retroviral recombination and reverse transcription. Science *250*, 1227-1233.

Hwang, S.S., Boyle, T.J., Lyerly, H.K., and Cullen, B.R. (1992). Identification of envelope V3 loop as the major determinant of CD4 neutralization sensitivity of HIV-1. Science *257*, 535-537.

Janssens, W., Nkengasong, J.N., Heyndrickx, L., Fransen, K., Ndumbe, P.M., Delaporte, E., Peeters, M., Perret, J.L., Ndoumou, A., Atende, C., and et al. (1994). Further evidence of the presence of genetically aberrant HIV-1 strains in Cameroon and Gabon. AIDS *8*, 1012-1013.

Jeang, K.T., Chun, R., Lin, N.H., Gatignol, A., Glabe, C.G., and Fan, H. (1993). In vitro and in vivo binding of human immunodeficiency virus type 1 Tat protein and Sp1 transcription factor. Journal of Virology *67*, 6224-6233.

Jones, K.A., Kadonaga, J.T., Luciw, P.A., and Tjian, R. (1986). Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. Science *232*, 755-759.

Jones, K.A., Luciw, P.A., and Duchange, N. (1988). Structural arrangements of transcription control domains within the 5'-untranslated leader regions of the HIV-1 and HIV-2 promoters. Genes & Development *2*, 1101-1114.

Jukes T.H., King J.L. (1979). Evolutionary nucleotide replacements in DNA. Nature 281, 605-606.

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

Kashanchi, F., Piras, G., Radonovich, M.F., Duvall, J.F., Fattaey, A., Chiang, C.M., Roeder, R.G., and Brady, J.N. (1994). Direct interaction of human TFIID with the HIV-1 transactivator tat. Nature *367*, 295-299.

Kellam, P., Boucher, C.A., Tijnagel, J.M., and Larder, B.A. (1994). Zidovudine treatment results in the selection of human immunodeficiency virus type 1 variants whose genotypes confer increasing levels of drug resistance. Journal of General Virology *75*, 341-351.

Kellam, P. and Larder, B.A. (1995). Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance. Journal of Virology *69*, 669-674.

Keohavong, P., Ling, L., Dias, C., and Thilly, W.G. (1993). Predominant mutations induced by the Thermococcus litoralis, vent DNA polymerase during DNA amplification in vitro. PCR Methods & Applications *2*, 288-292.

Keys, B., Karis, J., Fadeel, B., Valentin, A., Norkrans, G., Hagberg, L., and Chiodi, F. (1993). V3 Sequences of Paired HIV-1 Isolates from Blood and Cerebrospinal Fluid Cluster According to Host and Show Variation Related to the Clinical Stage of Disease. Virology *196*, 475-483.

Kido, H., Kamoshita, K., Fukutomi, A., and Katunuma, N. (1993). Processing protease for gp160 human immunodeficiency virus type I envelope glycoprotein precursor in human T4+ lymphocytes. Purification and characterization. Journal of Biological Chemistry *268*, 13406-13413.

Kim, S.Y., Byrn, R., Groopman, J., and Baltimore, D. (1989). Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. Journal of Virology *63*, 3708-3713.

Kinloch-De Loes, S., Hirschel, B.J., Hoen, B., Cooper, D.A., Tindall, B., Carr, A., Saurat, J.H., Clumeck, N., Lazzarin, A., Mathiesen, L., and et al. (1995). A controlled trial of zidovudine in primary human immunodeficiency virus infection [see comments]. New England Journal of Medicine *333*, 408-413.

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

Koenig, S., Gendelman, H.E., Orenstein, J.M., Dal Canto, M.C., Pezeshkpour, G.H., Yungbluth, M., Janotta, F., Aksamit, A., Martin, M.A., and Fauci, A.S. (1986). Detection of AIDS Virus in Macrophages in Brain Tissue from AIDS Patients with Encephalopathy. Science *233*, 1089-1093.

Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A., and Steitz, T.A. (1992). Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science *256*, 1783-1790.

Koken, S.E., van Wamel, J.L., Goudsmit, J., Berkhout, B., and Geelen, J.L. (1992). Natural variants of the HIV-1 long terminal repeat: analysis of promoters with duplicated DNA regulatory motifs. Virology *191*, 968-972.

Kollmus, H., Honigman, A., Panet, A., and Hauser, H. (1994). The sequences of and distance between two cis-acting signals determine the efficiency of ribosomal frameshifting in human immunodeficiency virus type 1 and human T-cell leukemia virus type II in vivo. Journal of Virology *68*, 6087-6091.

Koot, M., Keet, I.P., Vos, A.H., de Goede, R.E., Roos, M.T., Coutinho, R.A., Miedema, F, Schellekens, P.T., and Tersmette, M. (1993). Prognostic value of HIV-1 syncytium-inducing phenotype for rate of CD4+ cell depletion and progression to AIDS [see comments]. Annals of Internal Medicine *118*, 681-688.

Korber, B., Wolinsky, S., Haynes, B., Kunstman, K., Levy, R., Furtado, M., Otto, P, and Myers, G. (1992). HIV-1 intrapatient sequence diversity in the immunogenic V3 region. AIDS Research & Human Retroviruses *8*, 1461-1465.

Korber, B.T., MacInnes, K., Smith, R.F., and Myers, G. (1994). Mutational trends in V3 loop protein sequences observed in different genetic lineages of human immunodeficiency virus type 1. Journal of Virology *68*, 6730-6744.

Kostrikis, L.G., Bagdades, E., Cao, Y., Zhang, L., Dimitriou, D., and Ho, D.D. (1995). Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: Identification of a new subtype designated Subtype I. Journal of Virology *69*, 6122-6130.

Koup, R.A., Safrit, J.T., Cao, Y., Andrews, C.A., McLeod, G., Borkowsky, W., and Ho, D.D. (1994). Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. Journal of Virology *68*, 4650-4655.

Koyanagi, Y., Miles, S., Mitsuyasu, R.T., Merrill, J.E., Vinters, H.V., and Chen, I.S.Y. (1987). Dual Infection of the Central Nervous System by AIDS Viruses with Distinct Cellular Tropisms. Science *236*, 819-822.

Kuiken, C.L., de Jong, J.J., Baan, E., Keulen, W., Tersmette, M., and Goudsmit, J. (1992). Evolution of the V3 Envelope Domain in Proviral Sequences and Isolates of Human Immunodeficiency Virus Type 1 during Transition of the Viral Biological Phenotype. Journal of Virology *66*, 4622-4627.

Lafeuillade A, Tamalet C, de Micco Ph, and Thiebaut C (1995). Virological analysis of primary infection with rapid CD4 depletion. AIDS *9*, 1380-1381.

Lai, S., Page, J.B., and Lai, H. (1995). HIV results in the frame. Paradox remains . Nature *375*, 196-7.

Larder, B.A. (1994). Interactions between drug resistance mutations in human immunodeficiency virus type 1 reverse transcriptase. Journal of General Virology *75*, 951-957.

LaRosa, G.J., Davide, J.P., Weinhold, K., Waterbury, J.A., Profy, A.T., Lewis, J.A., Langlois, A.J., Dreesman, G.R., Boswell, R.N., Shadduck, P., Holley, L.H., Karplus, M., Bolognesi, D.P., Matthews, T.J., Emini, E.A., and Putney, S.D. (1990). Conserved Sequence and Structural Elements in the HIV-1 Principal Neutralizing Determinant. Science *249*, 932-935.

Laspia, M.F., Rice, A.P., and Mathews, M.B. (1989). HIV-1 Tat protein increases transcriptional initiation and stabilizes elongation. Cell *59*, 283-292.

Lazaro, I., Naniche, D., Signoret, N., Bernard, A.M., Marguet, D., Klatzmann, D., Dragic, T., Alizon, M., and Sattentau, Q. (1994). Factors involved in entry of the human immunodeficiency virus type 1 into permissive cells: lack of evidence of a role for CD26. Journal of Virology *68*, 6535-6546.

Lenz, J., Celander, D., Crowther, R.L., Patarca, R., Perkins, D.W., and Haseltine, W.A. (1984). Determination of the leukaemogenicity of a murine retrovirus by sequences within the long terminal repeat. Nature *308*, 467-470.

Leonard, J., Parrott, C., Buckler-White, A.J., Turner, W., Ross, E.K., Martin, M.A., and Rabson, A.B. (1989). The NF-kappa B binding sites in the human immunodeficiency virus type 1 long terminal repeat are not required for virus infectivity. Journal of Virology *63*, 4919-4924.

Levy, J.A. (1994). HIV and the Pathogenesis of AIDS. ASM Press, Washington, D.C., USA.

Li, Y.C., Ross, J., Scheppler, J.A., and Franza, B.R., Jr. (1991). An in vitro transcription analysis of early responses of the human immunodeficiency virus type 1 long terminal repeat to different transcriptional activators. Molecular & Cellular Biology *11*, 1883-1893.

Ling, L.L., Keohavong, P., Dias, C., and Thilly, W.G. (1991). Optimization of the polymerase chain reaction with regard to fidelity: modified T7, Taq, and vent DNA polymerases. PCR Methods & Applications *1*, 63-69.

Linial, M.L. and Miller, A.D. (1990). Retroviral RNA packaging: sequence requirements and implications. Current Topics in Microbiology & Immunology *157*, 125-152.

Loussert-Ajaka, I., Ly, T.D., Chaix, M.L., Ingrand, D., Saragosti, S., Courouce, AM, Brun-Vezinet, F., and Simon, F. (1994). HIV-1/HIV-2 seronegativity in HIV-1 subtype O infected patients. Lancet *343*, 1393-1394.

Loussert-Ajaka, I., Chaix, M.L., Korber, B., Letourneur, F., Gomas, E., Allen, E., Ly, T.D., Brun-Vezinet, F., Simon, F., and Saragosti, S. (1995). Variability of human immunodeficiency virus type 1 group O strains isolated from Cameroonian patients living in France. Journal of Virology *69*, 5640-5649.

Louwagie, J., Janssens, W., Mascola, J., Heyndrickx, L., Hegerich, P., van der Groen, G., McCutchan, F.E., and Burke, D.S. (1995). Genetic Diversity of the Envelope Glycoprotein from Human Immunodeficiency Virus Type 1 Isolates of African Origin. Journal of Virology *69*, 263-271.

Lu, Y., Stenzel, M., Sodroski, J.G., and Haseltine, W.A. (1989). Effects of long terminal repeat mutations on human immunodeficiency virus type 1 replication. Journal of Virology *63*, 4115-4119.

Lu, Y.C., Touzjian, N., Stenzel, M., Dorfman, T., Sodroski, J.G., and Haseltine, W.A. (1990). Identification of cis-acting repressive sequences within the negative regulatory element of human immunodeficiency virus type 1. Journal of Virology *64*, 5226-5229.

Lundberg, K.S., Shoemaker, D.D., Adams, M.W., Short, J.M., Sorge, J.A., and Mathur, E.J. (1991). High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus. Gene *108*, 1-6.

Luo, Y., Yu, H., and Peterlin, B.M. (1994). Cellular protein modulates effects of human immunodeficiency virus type 1 Rev. Journal of Virology *68*, 3850-3856.

Maddon, P.J., Dalgleish, A.G., McDougal, J.S., Clapham, P.R., Weiss, R.A., and Axel, R. (1986). The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. Cell *47*, 333-348.

Maddon, P.J., McDougal, J.S., Clapham, P.R., Dalgleish, A.G., Jamal, S., Weiss, R.A., and Axel, R. (1988). HIV infection does not require endocytosis of its receptor, CD4. Cell *54*, 865-874.

Madore, S.J., Tiley, L.S., Malim, M.H., and Cullen, B.R. (1994). Sequence requirements for Rev multimerization in vivo. Virology *202*, 186-194.

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

Malim, M.H. and Cullen, B.R. (1991). HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency. Cell *65*, 241-248.

Mansky, L.M. and Temin, H.M. (1995). Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. Journal of Virology *69*, 5087-5094.

Marciniak, R.A., Calnan, B.J., Frankel, A.D., and Sharp, P.A. (1990a). HIV-1 Tat protein trans-activates transcription in vitro. Cell *63*, 791-802.

Marciniak, R.A., Garcia-Blanco, M.A., and Sharp, P.A. (1990b). Identification and characterization of a HeLa nuclear protein that specifically binds to the trans-activation-response (TAR) element of human immunodeficiency virus. Proceedings of the National Academy of Sciences, USA *87*, 3624-3628.

McCune, J.M., Rabin, L.B., Feinberg, M.B., Lieberman, M., Kosek, J.C., Reyes, G.R., and Weissman, I.L. (1988). Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus. Cell *53*, 55-67.

McCune, J.M. (1991). HIV-1: the infective process in vivo. Cell *64*, 351-363.

McKnight, A., Weiss, R.A., Shotton, C., Takeuchi, Y., Hoshino, H., and Clapham, P.R. (1995). Change in tropism upon immune escape by human immunodeficiency virus. Journal of Virology *69*, 3167-3170.

McLean, A.R. and Nowak, M.A. (1992). Competition between zidovudine-sensitive and zidovudine-resistant strains of HIV. AIDS *6*, 71-79.

McNearney T, Hornickova Z, Templeton A, Bidrwell A, Arens M, Markham R, Saah A, and Ratner L (1995). Nef and LTR Sequence Variation from Sequentially Derived Human Immunodeficiency Virus Type 1 isolates. Virology *208*, 388-398.

McNearney, T., Hornickova, Z., Markham, R., Birdwell, A., Arens, M., Saah, A., and Ratner, L. (1992). Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. Proceedings of the National Academy of Sciences. *89*, 10247-10251.

Mellors, J.W., Kingsley, L.A., Rinaldo, C.R., Jr., Todd, J.A., Hoo, B.S., Kokka, R.P., and Gupta, P. (1995). Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. Annals of Internal Medicine *122*, 573-579.

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

Meyerhans, A., Dadaglio, G., Vartanian, J.P., Langlade-Demoyen, P., Frank, R., Plata, F., and Wain-Hobson, S. (1991). In vivo persistence of a HIV-1-encoded HLA-B27-restricted cytotoxic T lymphocyte epitope despite specific in vitro reactivity. European Journal of Immunology *21*, 2637-2640.

Michael, N.L., D'Arcy, L., Ehrenberg, P.K., and Redfield, R.R. (1994). Naturally occurring genotypes of the human immunodeficiency virus type 1 long terminal repeat display a wide range of basal and Tat-induced transcriptional activities. Journal of Virology *68*, 3163-3174.

Michie, C. (1995). HIV results in the frame. Toxic shock. Nature *375*, 197-8.

Milich, L., Margolin, B., and Swanstrom, R. (1993). V3 loop of the human immunodeficiency virus type 1 Env protein: interpreting sequence variability. Journal of Virology *67*, 5623-5634.

Moore, P.D., Bose, K.K., Rabkin, S.D., and Strauss, B.S. (1981). Sites of termination of in vitro DNA synthesis on ultraviolet-and N-acetylaminofluorene-treated phi X174 templates by prokaryotic and eukaryotic DNA polymerases. Proceedings of the National Academy, USA *78*, 110-114.

Morikawa, Y., Kishi, T., Zhang, W.H., Nermut, M.V., Hockley, D.J., and Jones, I.M. (1995). A molecular determinant of human immunodeficiency virus particle assembly located in matrix antigen p17. Journal of Virology *69*, 4519-4523.

Moses, A.V., Ibanez, C., Gaynor, R., Ghazal, P., and Nelson, J.A. (1994). Differential role of long terminal repeat control elements for the regulation of basal and Tat-mediated transcription of the human immunodeficiency virus in stimulated and unstimulated primary human macrophages. Journal of Virology *68*, 298-307.

Mosier, D.E. (1995). HIV results in the frame. CD4+ cell turnover. Nature *375*, 193-4.

Mulder-Kampinga, G.A., Simonon, A., Kuiken, C.L., Dekker, J., Scherpbier, H.J., van de Perre, P., Boer, K., and Goudsmit, J. (1995). Similarity in env and gag genes between genomic RNAs of human immunodeficiency virus type 1 (HIV-1) from mother and infant is unrelated to time of HIV-1 RNA positivity in the child. Journal of Virology *69*, 2285-2296.

Myers, G. (1994). Tenth anniversary perspectives on AIDS. HIV: between past and future. AIDS Research & Human Retroviruses *10*, 1317-1324.

Nabel, G.J. (1988). Activation of human immunodeficiency virus. Journal of Laboratory & Clinical Medicine *111*, 495-500.

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

Nagai, Y. (1993). Protease-dependent virus tropism and pathogenicity. Trends in Microbiology *1*, 81-87.

Nakanishi, Y., Masamune, Y., and Kobayashi, N. (1991). A novel cis-acting element that controls transcription of human immunodeficiency virus type 1 DNA, depending on cell type. Journal of Virology *65*, 6334-6338.

Newstein, M., Stanbridge, E.J., Casey, G., and Shank, P.R. (1990). Human chromosome 12 encodes a species-specific factor which increases human immunodeficiency virus type 1 tat-mediated trans activation in rodent cells. Journal of Virology *64*, 4565-4567.

Nikolov, D.B., Hu, S.H., Lin, J., Gasch, A., Hoffmann, A., Horikoshi, M., Chua, N.H., Roeder, R.G., and Burley, S.K. (1992). Crystal structure of TFIID TATA-box binding protein. Nature *360*, 40-46.

Nikolov, D.B. and Burley, S.K. (1994). 2.1 A resolution refined structure of a TATA box-binding protein (TBP). Nature Structural Biology *1*, 621-637.

Nkengasong, J.N., Janssens, W., Heyndrickx, L., Fransen, K., Ndumbe, P.M., Motte, J, Leonaers, A., Ngolle, M., Ayuk, J., Piot, P., and et al. (1994). Genotypic subtypes of HIV-1 in Cameroon. AIDS *8*, 1405-1412.

Northrop, J.P., Ho, S.N., Chen, L., Thomas, D.J., Timmerman, L.A., Nolan, G.P., AdmonA., and Crabtree, G.R. (1994). NF-AT components define a family of transcription factors targeted in T-cell activation. Nature *369*, 497-502.

Nowak, M.A., Bonhoeffer, S., Loveday, C., Balfe, P., Semple, M., Kaye, S., Tenant-Flowers, M., and Tedder, R. (1995a). HIV results in the frame. Results confirmed [letter; comment]. Nature *375*, 193

Nowak, M.A., May, R.M., Phillips, R.E., Rowland-Jones, S., Lalloo, D.G., McAdam, S., Klenerman, P., Koppe, B., Sigmund, K., Bangham, C.R., and McMichael A. (1995b). Antigenic oscillations and shifting immunodominance in HIV-1 infections. Nature *375*, 606-611.

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

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

O'Brien, W.A., Chen, I.S., Ho, D.D., and Daar, E.S. (1992). Mapping genetic determinants for human immunodeficiency virus type 1 resistance to soluble CD4. Journal of Virology *66*, 3125-3130.

Ohnishi, Y., Shioda, T., Nakayama, K., Iwata, S., Gotoh, B., Hamaguchi, M., and Nagai, Y. (1994). A furin-defective cell line is able to process correctly the gp160 of human immunodeficiency virus type 1. Journal of Virology *68*, 4075-4079.

Orchard, K., Perkins, N., Chapman, C., Harris, J., Emery, V., Goodwin, G., Latchman, D., and Collins, M. (1990). A T-cell protein which recognizes a palindromic DNA sequence in the negative regulatory element of the HIV-1 long terminal repeat with homology to steroid/thyroid hormone receptor binding sites. Biochemical Society Transactions *18*, 555-556.

Overbeek, P.A., Lai, S.P., Van Quill, K.R., and Westphal, H. (1986). Tissue-specific expression in transgenic mice of a fused gene containing RSV terminal sequences. Science *231*, 1574-1577.

Palker, T.J., Clark, M.E., Langlois, A.J., Matthews, T.J., Weinhold, K.J., Randall, RR, Bolognesi, D.P., and Haynes, B.F. (1988). Type-specific neutralization of the human immunodeficiency virus with antibodies to env-encoded synthetic peptides. Proceedings of the National Academy of Sciences, USA *85*, 1932-1936.

Pang, S., Koyanagi, Y., Miles, S., Wiley, C., Vinters, H.V., and Chen, I.S. (1990). High levels of unintegrated HIV-1 DNA in brain tissue of AIDS dementia patients. Nature *343*, 85-89.

Pang, S., Vinters, H.V., Akashi, T., O'Brien, W.A., and Chen, I.S. (1991). HIV-1 env sequence variation in brain tissue of patients with AIDS-related neurologic disease. Journal of Acquired Immune Deficiency Syndromes *4*, 1082-1092.

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

Pantaleo, G., Graziosi, C., Butini, L., Pizzo, P.A., Schnittman, S.M., Kotler, D.P., and Fauci, A.S. (1991). Lymphoid organs function as major reservoirs for human immunodeficiency virus. Proceedings of the National Academy of Sciences, USA *88*, 9838-9842.

Pantaleo, G., Graziosi, C., Demarest, J.F., Butini, L., Montroni, M., Fox, C.H., Orenstein, J.M., Kotler, D.P., and Fauci, A.S. (1993a). HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease. Nature *362*, 355-358.

Pantaleo, G., Graziosi, C., and Fauci, A.S. (1993b). The role of lymphoid organs in the immunopathogenesis of HIV infection. AIDS *7(supp1)*, S19-S23.

Pantaleo, G., Demarest, J.F., Soudeyns, H., Graziosi, C., Denis, F., Adelsberger, JW, Borrow, P., Saag, M.S., Shaw, G.M., Sekaly, R.P., and et al (1994). Major expansion of CD8+ T cells with a predominant V beta usage during the primary immune response to HIV. Nature *370*, 463-467.

Pantaleo, G., Menzo, S., Vaccarezza, M., Graziosi, C., Cohen, O.J., Demarest, J.F., Montefiori, D., Orenstein, J.M., Fox, C., Schrager, L.K., Margolick, J.B., Buchbinder, S., Giorgi, J.V., and Fauci, A.S. (1995). Studies in Subjects with Long-term Nonprogressive Human Immunodeficiency Virus Infection. New England Journal of Medicine *332 (4)*, 209-216.

Parrott, C., Seidner, T., Duh, E., Leonard, J., Theodore, T.S., Buckler-White, A., Martin, M.A., and Rabson, A.B. (1991). Variable role of the long terminal repeat Sp1-binding sites in human immunodeficiency virus replication in T lymphocytes. Journal of Virology *65*, 1414-1419.

Patel, P.H. and Preston, B.D. (1994). Marked infidelity of human immunodeficiency virus type 1 reverse transcriptase at RNA and DNA template ends. Proceedings of the National Academy of Sciences, USA *91*, 549-553.

Patience, C., McKnight, A., Clapham, P.R., Boyd, M.T., Weiss, R.A., and Schulz, T.F. (1994). CD26 antigen and HIV fusion?. Science *264*, 1159-60.

Pedroza Martins, L., Chenciner, N., and Wain-Hobson, S. (1992). Complex intrapatient sequence variation in the V1 and V2 hypervariable regions of the HIV-1 gp 120 envelope sequence. Virology *191*, 837-845.

Phillips, A.N., Sabin, C.A., Mocroft, A., and Janossy, G. (1995). HIV results in the frame. Antiviral therapy. Nature *375*, 195.

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

Pillay, D., Lipman, M.C., Lee, C.A., Johnson, M.A., Griffiths, P.D., and McLaughlin, J.E. (1993). A clinico-pathological audit of opportunistic viral infections in HIV-infected patients. AIDS *7*, 969-974.

Power, C., McArthur, J.C., Johnson, R.T., Griffin, D.E., Glass, J.D., Perryman, S., and Chesebro, B. (1994). Demented and Nondemented Patients with AIDS Differ in Brain-Derived Human Immunodeficiency Virus Type 1 Envelope Sequences. Journal of Virology *68*, 4643-4649.

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

Price, R.W., Brew, B., Sidtis, J., Rosenblum, M., Scheck, A.C., and Cleary, P. (1988). The brain in AIDS: central nervous system HIV-1 infection and AIDS dementia complex. Science *239*, 586-592.

Ratnasabapathy, R., Sheldon, M., Johal, L., and Hernandez, N. (1990). The HIV-1 long terminal repeat contains an unusual element that induces the synthesis of short RNAs from various mRNA and snRNA promoters. Genes & Development *4*, 2061-2074.

Reil, H., Kollmus, H., Weidle, U.H., and Hauser, H. (1993). A heptanucleotide sequence mediates ribosomal frameshifting in mammalian cells. Journal of Virology *67*, 5579-5584.

Reitz, M.S., Jr., Wilson, C., Naugle, C., Gallo, R.C., and Robert-Guroff, M. (1988). Generation of a neutralization-resistant variant of HIV-1 is due to selection for a point mutation in the envelope gene. Cell *54*, 57-63.

Richman, D.D. and Bozzette, S.A. (1994). The impact of the syncytium-inducing phenotype of human immunodeficiency virus on disease progression. Journal of Infectious Diseases *169*, 968-974.

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

Rosen, C.A., Haseltine, W.A., Lenz, J., Ruprecht, R., and Cloyd, M.W. (1985a). Tissue selectivity of murine leukemia virus infection is determined by long terminal repeat sequences. Journal of Virology *55*, 862-866.

Rosen, C.A., Sodroski, J.G., and Haseltine, W.A. (1985b). 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.

Rothblum, C.J., Jackman, J., Mikovits, J., Shukla, R.R., and Kumar, A. (1995). Interaction of nuclear protein p140 with human immunodeficiency virus type 1 TAR RNA in mitogen-activated primary human T lymphocytes. Journal of Virology *69*, 5156-5163.

Rounseville, M.P. and Kumar, A. (1992). Binding of a host cell nuclear protein to the stem region of human immunodeficiency virus type 1 trans-activation-responsive RNA. Journal of Virology *66*, 1688-1694.

Rowland-Jones, S.L., Phillips, R.E., Nixon, D.F., Gotch, F.M., Edwards, J.P., Ogunlesi, A.O., Elvin, J.G., Rothbard, J.A., Bangham, C.R., Rizza, C.R., and et al. (1992). Human immunodeficiency virus variants that escape cytotoxic T-cell recognition. AIDS Research & Human Retroviruses *8*, 1353-1354.

Roy, S., Delling, U., Chen, C.H., Rosen, C.A., and Sonenberg, N. (1990). A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. Genes & Development *4*, 1365-1373.

Rusche, J.R., Javaherian, K., McDanal, C., Petro, J., Lynn, D.L., Grimaila, R., Langlois, A., Gallo, R.C., Arthur, L.O., Fischinger, P.J., 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. Proceedings of the National Academy of Sciences, USA *85*, 3198-3202.

Safrit J.T. and Koup R.A. (1996). The CD4 Loss in AIDS Patients is *not* Immunopathologically Mediated. Reviews in Medical Virology (in press).

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

Sambrook J., Fritsch E.F., and Maniatis T. (1989). Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory Press.

Sarkar, G. and Sommer, S.S. (1990). Shedding light on PCR contamination. Nature *343*, 27

Scarlatti, G., Leitner, T., Halapi, E., Wahlberg, J., Marchisio, P., Clerici-Schoeller, M.A., Wigzell, H., Fenyo, E.M., Albert, J., Uhlen, M., and et al. (1993). Comparison of variable region 3 sequences of human immunodeficiency virus type 1 from infected children with the RNA and DNA sequences of the virus populations of their mothers. Proceedings of the National Academy of Sciences, USA *90*, 1721-1725.

Sharp, P.M., Robertson, D.L., Gao, F., and Hahn, B.H. (1994). Origins and diversity of human immunodeficiency viruses. AIDS *8 (supp1)*, S27-S42.

Sharpless, N.E., O'Brien, W.A., Verdin, E., Kufta, C.V., Chen, I.S., Dubois-Dalcq, and M. (1992). Human immunodeficiency virus type 1 tropism for brain microglial cells is determined by a region of the env glycoprotein that also controls macrophage tropism. Journal of Virology *66*, 2588-2593.

Signoret, N., Poignard, P., Blanc, D., and Sattentau, Q.J. (1993). Human and simian immunodeficiency viruses: virus-receptor interactions. Trends in Microbiology *1*, 328-333.

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

Simmonds, P., Zhang, L.Q., McOmish, F., Balfe, P., Ludlam, C.A., and Brown, A.J. (1991). Discontinuous sequence change of human immunodeficiency virus (HIV) type 1 env sequences in plasma viral and lymphocyte-associated proviral populations in vivo: implications for models of HIV pathogenesis. Journal of Virology *65*, 6266-6276.

Small, J.A., Bieberich, C., Ghotbi, Z., Hess, J., Scangos, G.A., and Clements, J.E. (1989). The visna virus long terminal repeat directs expression of a reporter gene in activated macrophages, lymphocytes, and the central nervous systems of transgenic mice. Journal of Virology *63*, 1891-1896.

Southgate, C., Zapp, M.L., and Green, M.R. (1990). Activation of transcription by HIV-1 Tat protein tethered to nascent RNA through another protein. Nature *345*, 640-642.

Sprent, J. and Tough, D. (1995). HIV results in the frame. CD4+ cell turnover. Nature *375*, 194.

Squire, S.B., Lipman, M.C., Bagdades, E.K., Mulvenna, P.M., Grundy, J.E., GriffithsPD., and Johnson, M.A. (1992). Severe cytomegalovirus pneumonitis in HIV infected patients with higher than average CD4 counts. Thorax *47*, 301-304.

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

Sune, C. and Garcia-Blanco, M.A. (1995). Sp1 Transcription Factor Is Required for In Vitro Basal and Tat-Activated Transcription from the Human Immunodeficiency Virus Type 1 Long Terminal Repeat. Journal of Virology *69*, 6572-6576.

Takeuchi, Y., Akutsu, M., Murayama, K., Shimizu, N., and Hoshino, H. (1991). Host Range Mutant of Human Immunodeficiency Virus Type 1: Modification of Cell Tropism by a Single Point Mutation at the Neutralization Epitope in the env Gene. Journal of Virology *65*, 1710-1718.

Tamalet, C., Lafeuillade, A., Yahi, N., Vignoli, C., Tourres, C., Pellegrino, P., and de Micco, P. (1994). Comparison of viral burden and phenotype of HIV-1 isolates from lymph nodes and blood. AIDS *8*, 1083-1088.

Taylor, J.P., Pomerantz, R., Bagasra, O., Chowdhury, M., Rappaport, J., KhaliliK., and Amini, S. (1992). TAR-independent transactivation by Tat in cells derived from the CNS: a novel mechanism of HIV-1 gene regulation. EMBO Journal *11*, 3395-3403.

Taylor, J.P., Pomerantz, R.J., Raj, G.V., Kashanchi, F., Brady, J.N., Amini, S., and Khalili, K. (1994). Central nervous system-derived cells express a kappa B-binding activity that enhances human immunodeficiency virus type 1 transcription in vitro and facilitates TAR-independent transactivation by Tat. Journal of Virology *68*, 3971-3981.

Tenner-Racz, K., Racz, P., Bofill, M., Schulz-Meyer, A., Dietrich, M., Kern, P., Weber, J., Pinching, A.J., Veronese-Dimarzo, F., Popovic, M., Klatzmann, D., Gluckman, J.C., and Janossy, G. (1986). HTLV-III/LAV Viral Antigens in Lymph Nodes of Homosexual Men With Persistent Generalized Lymphadenopathy and AIDS. American Journal of Pathology *123(1)*, 9-15.

Tersmette, M., de Goede, R.E., Al, B.J., Winkel, I.N., Gruters, R.A., Cuypers, H.T., Huisman, H.G., and Miedema, F. (1988). Differential syncytium-inducing capacity of human immunodeficiency virus isolates: frequent detection of syncytium-inducing isolates in patients with acquired immunodeficiency syndrome (AIDS) and AIDS-related complex. Journal of Virology *62*, 2026-2032.

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

Tersmette, M., Lange, J.M., de Goede, R.E., de Wolf, F., Eeftink-Schattenkerk, JK, Schellekens, P.T., Coutinho, R.A., Huisman, J.G., Goudsmit, J., and Miedema, F. (1989b). Association between biological properties of human immunodeficiency virus variants and risk for AIDS and AIDS mortality. Lancet *1*, 983-985.

Tidy, J.A., Vousden, K.H., and Farrell, P.J. (1989). Relation between infection with a subtype of HPV16 and cervical neoplasia [published erratum appears in Lancet 1989 Oct 21; 2(8669):996] [retracted by Tidy J and Farrell PJ. In: Lancet 1989 Dec 23-30;2(8678-8679):1535]. Lancet *1*, 1225-1227.

Toggas, S.M., Masliah, E., Rockenstein, E.M., Rall, G.F., Abraham, C.R., and Mucke, L. (1994). Central nervous system damage produced by expression of the HIV-1 coat protein gp120 in transgenic mice. Nature *367*, 188-193.

Tsang, M.L., Evans, L.A., McQueen, P., Hurren, L., Byrne, C., Penny, R., Tindall, B., and Cooper, D.A. (1994). Neutralizing Antibodies against Sequential Autologous Human Immunodeficiency Virus Type 1 Isolates after Seroconversion. Journal of Infectious Diseases *170*, 1141-1147.

Vazeux, R., Brousse, N., Jarry, A., Henin, D., Marche, C., Vedrenne, C., Mikol, J, Wolff, M., Michon, C., Rozenbaum, W., and et al. (1987). AIDS subacute encephalitis. Identification of HIV-infected cells. American Journal of Pathology *126*, 403-410.

Wain-Hobson, S. (1993). The fastest genome evolution ever described: HIV variation in situ. [Review]. Current Opinion in Genetics & Development *3*, 878-883.

Wain-Hobson, S. (1995). Virological mayhem. Nature *373*, 102

Wainberg M.A. and Gu Z. (1995). Targeting HIV reverse transcriptase in novel ways. Nature Medicine *1*, 628-629.

Watkins, B.A., Dorn, H.H., Kelly, W.B., Armstrong, R.C., Potts, B.J., Michaels, F., Kufta, C.V., and Dubois-Dalcq, M. (1990). Specific Tropism of HIV-1 for Microglial Cells in Primary Human Brain Cultures. Science *249*, 549-553.

Watkins, B.A., Reitz, M.S., Jr., Wilson, C.A., Aldrich, K., Davis, A.E., and Robert-Guroff, M. (1993). Immune escape by human immunodeficiency virus type 1 from neutralizing antibodies: evidence for multiple pathways. Journal of Virology *67*, 7493-7500.

Weber, J. and Galpin, S. (1995). HIV results in the frame. Cyclosporin A. Nature *375*, 198.

Wei, X., Ghosh, S.K., Taylor, M.E., Johnson, V.A., Emini, E.A., Deutsch, P., Lifson, J.D., Bonhoeffer, S., Nowak, M.A., Hahn, B.H., Saag, M.S., and Shaw, G.M. (1995). Viral dynamics in human immunodeficiency virus type 1 infection. Nature *373*, 117-122.

Werner, A. and Levy, J.A. (1993). Human immunodeficiency virus type 1 envelope gp120 is cleaved after incubation with recombinant soluble CD4. Journal of Virology *67*, 2566-2574.

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

Zawel, L. and Reinberg, D. (1992). Advances in RNA polymerase II transcription. Current Opinion in Cell Biology *4*, 488-495.

Zekeng, L., Gurtler, L., Afane Ze, E., Sam-Abbenyi, A., Mbouni-Essomba, G., Mpoudi-Ngolle, E., Monny-Lobe, M., Tapka, J.B., and Kaptue, L. (1994). Prevalence of HIV-1 subtype O infection in Cameroon: preliminary results. AIDS *8*, 1626-1628.

Zhang, L.Q., MacKenzie, P., Cleland, A., Holmes, E.C., Brown, A.J., and Simmonds, P. (1993). Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. Journal of Virology *67*, 3345-3356.

Zhu, T., Mo, H., Wang, N., Nam, D.S., Cao, Y., Koup, R.A., and Ho, D.D. (1993). Genotypic and Phenotypic Characterization of HIV-1 in Patients with Primary Infection. Science *261*, 1179-1181.

Zinkernagel, R.M. and Hengartner, H. (1994). T-cell-mediated immunopathology versus direct cytolysis by virus: implications for HIV and AIDS. Immunology Today *15*, 262-268.

Zvi, A., Kustanovich, I., Feigelson, D., Levy, R., Eisenstein, M., Matsushita, S, Richalet-Secordel, P., Regenmortel, M.H., and Anglister, J. (1995a). NMR mapping of the antigenic determinant recognized by an anti-gp120, human immunodeficiency virus neutralizing antibody. European Journal of Biochemistry *229*, 178-187.

Zvi, A., Kustanovich, I., Hayek, Y., Matsushita, S., and Anglister, J. (1995b). The principal neutralizing determinant of HIV-1 located in V3 of gp120 forms a 12-residue loop by internal hydrophobic interactions. FEBS Letters *368*, 267-270.