GENETIC ANALYSIS OF POPULATIONS OF HIV-1 VARIANTS INFECTING DIFFERENT TISSUES in vivo.

ELIZABETH.S.HUGHES

Thesis presented for the degree of Doctor of Philosophy.

The University of Edinburgh.

1997.
DECLARATION

All of the procedures and investigations described in this thesis have been performed by the author.

The contents of this thesis were composed by the author.
ACKNOWLEDGEMENTS

I would like to take this opportunity to thank Dr Myra Arnott, my colleagues from the Molecular Virology Laboratory and the staff of the Hepatitis Reference Laboratory for all their help and support. I would also like to thank my supervisors Dr Pete Simmonds and Dr Jeanne Bell for their unending efforts in guiding me through my PhD.

My thanks also go to my family and close friends without whom I would not have made it through this experience sane.

This study was funded by an MRC ADP studentship grant.
Infection with human immunodeficiency type 1 (HIV-1) is associated with a slow irreversible impairment of the immune system eventually leading to AIDS. Many studies have been carried out to elucidate the infection mechanisms and pathogenicity of HIV-1 although a number of questions still remain. This study was carried out to investigate the relationship between HIV-1 populations infecting different tissues to determine when infection to non-lymphoid tissue occurs during the course of infection.

The time of spread of HIV-1 to non-lymphoid tissue was investigated by sequence comparisons of variants infecting a range of tissues from three individuals with AIDS in the p17 gag gene, and flanking regions of V1/V2. In both regions, phylogenetic analysis revealed several lineages in each individual that contained sequences from lymphoid and non-lymphoid tissue, such as brain. This observation contrasted strongly with the previously described organ-specific sequences in the V3 region in this study population and in other investigations.

By estimating mean synonymous pairwise distances in the p17 gag region, it was possible to calculate the time of divergence of variants infecting lymphoid and non-lymphoid tissues. In lymphoid tissue the mean diversity of gag sequences implied an approximate population age of 2.65 to 5.6 years, while those infecting brain were significantly more variable, suggesting an even earlier time of diversification (4.1 to 6.2 years). In two of the three individuals, these times of divergence indicate that infection of the brain may have occurred as an early event in the progression of the disease, preceding the onset of AIDS by several years.
This is the first report in which it has been possible to estimate times of diversification in different tissues \textit{in vivo} and is of importance in understanding the dynamics of the spread of HIV-1 into non-lymphoid tissue, and its possible adaptation for replication in different cell types.

Phenotypic variation between different isolates of HIV-1, such as macrophage tropism and syncytium induction have been mapped to specific regions in the \textit{env} gene, including the V3 and V1/V2 hypervariable regions. In this study I have analysed variability in the V1 and V2 regions of HIV-1 proviral sequences amplified from lymphoid tissue, brain and other non-lymphoid tissue collected at autopsy from three HIV-infected individuals with giant cell encephalitis. As previously found by analysing p17\textsubscript{gag} region I found no evidence for any tissue-specific grouping of variants in the V1/V2 regions. Furthermore, I found no correlation of charge, length or number of glycosylation sites with tissue origin, or inferred phenotype.

Length polymorphism analysis is a rapid method to compare whole populations of HIV-1 variants within a sample, and provides information on the length and diversity of the V1 and V2 hypervariable regions. Based upon a comparison of 42 individuals with CD4 counts ranging from 802 to <1, at time of death, I found no evidence for changes in the length of V2. Using the number of length variants in the V1 and V2 hypervariable region as a measure of the overall degree of variability within HIV populations, I found no evidence for an increase or a decrease in diversity between those with or without AIDS defining illness.

In summary, both analysis of the p17\textsubscript{gag} and V1/V2 regions revealed high levels of heterogeneity of HIV-1 in tissue such as brain producing multiple lineages.
upon phylogenetic analysis. I have therefore found no evidence for specifically neurotropic variants of HIV-1 and question the idea that spread into the central nervous system (CNS) or other non-lymphoid tissues requires specific adaptation.
<table>
<thead>
<tr>
<th>ABBREVIATIONS</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADCC</td>
<td>antibody dependent cellular cytotoxicity</td>
</tr>
<tr>
<td>ADC</td>
<td>AIDS dementia complex</td>
</tr>
<tr>
<td>ADP</td>
<td>AIDS directed programme</td>
</tr>
<tr>
<td>AIDS</td>
<td>acquired immune deficiency syndrome</td>
</tr>
<tr>
<td>ALV</td>
<td>avian leukaemia virus</td>
</tr>
<tr>
<td>ARV</td>
<td>AIDS related virus</td>
</tr>
<tr>
<td>AZT</td>
<td>3'-azido-3-deoxythymidine (zidovudine)</td>
</tr>
<tr>
<td>β-2-M</td>
<td>beta-2-microglobulin</td>
</tr>
<tr>
<td>BIV</td>
<td>bovine immunodeficiency virus</td>
</tr>
<tr>
<td>BMVEC</td>
<td>brain microvascular endothelial cells</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>BW</td>
<td>binding and washing buffer</td>
</tr>
<tr>
<td>Ca²⁺</td>
<td>calcium ion</td>
</tr>
<tr>
<td>CAT</td>
<td>chloramphenicol acetyltransferase</td>
</tr>
<tr>
<td>CD</td>
<td>cluster determinant</td>
</tr>
<tr>
<td>CDC</td>
<td>Centre for Disease Control</td>
</tr>
<tr>
<td>CDR</td>
<td>complementary determining region</td>
</tr>
<tr>
<td>CHO</td>
<td>glycosylation</td>
</tr>
<tr>
<td>CMI</td>
<td>cell mediated immunity</td>
</tr>
<tr>
<td>CNS</td>
<td>central nervous system</td>
</tr>
<tr>
<td>CSF</td>
<td>cerebrospinal fluid</td>
</tr>
<tr>
<td>CTL</td>
<td>cytotoxic T lymphocyte</td>
</tr>
<tr>
<td>CTLp</td>
<td>cytotoxic T lymphocyte precursor</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>dATP</td>
<td>deoxyadenosine triphosphate</td>
</tr>
<tr>
<td>ddC</td>
<td>2',3'-dideoxycytidine (zalcitabine)</td>
</tr>
<tr>
<td>ddI</td>
<td>2',3'-dideoxyinosine</td>
</tr>
<tr>
<td>d4T</td>
<td>stavudine</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethylsulphoxide</td>
</tr>
<tr>
<td>$d_N$</td>
<td>rate of nonsynonymous substitutions</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNA dep RNA pol</td>
<td>DNA dependent RNA polymerase</td>
</tr>
<tr>
<td>ds</td>
<td>double stranded</td>
</tr>
<tr>
<td>$d_s$</td>
<td>rate of synonymous substitutions</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxyribonucleotides</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EIA</td>
<td>enzyme immunoassay</td>
</tr>
<tr>
<td>EM</td>
<td>electron microscopic</td>
</tr>
<tr>
<td>FeLV</td>
<td>feline leukaemia virus</td>
</tr>
<tr>
<td>FIV</td>
<td>feline immunodeficiency virus</td>
</tr>
<tr>
<td>FMDV</td>
<td>foot and mouth disease</td>
</tr>
<tr>
<td>GCE</td>
<td>giant cell encephalitis</td>
</tr>
<tr>
<td>GM-CSF</td>
<td>granulocyte macrophage colony stimulating factor</td>
</tr>
<tr>
<td>gp</td>
<td>glycoprotein</td>
</tr>
<tr>
<td>HIV</td>
<td>human immunodeficiency virus</td>
</tr>
<tr>
<td>HTLV</td>
<td>human T-cell leukaemia virus</td>
</tr>
<tr>
<td>IDU</td>
<td>intravenous drug users</td>
</tr>
<tr>
<td>IFN</td>
<td>interferon</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>Ig</td>
<td>immunoglobulin</td>
</tr>
<tr>
<td>IL</td>
<td>interleukin</td>
</tr>
<tr>
<td>IN</td>
<td>integrase</td>
</tr>
<tr>
<td>Kb</td>
<td>kilobases</td>
</tr>
<tr>
<td>Kd</td>
<td>kilodalton</td>
</tr>
<tr>
<td>LAV</td>
<td>lymphadenopathy associated virus</td>
</tr>
<tr>
<td>LCMV</td>
<td>lymphocytic choriomeningitis virus</td>
</tr>
<tr>
<td>LFA-1</td>
<td>leucocyte function accessory-1 molecule</td>
</tr>
<tr>
<td>LPA</td>
<td>length polymorphism analysis</td>
</tr>
<tr>
<td>LTR</td>
<td>long terminal repeat</td>
</tr>
<tr>
<td>MA</td>
<td>matrix</td>
</tr>
<tr>
<td>MGC</td>
<td>multinucleated giant cell</td>
</tr>
<tr>
<td>MHC</td>
<td>major histocompatibility complex</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>MuLV</td>
<td>murine leukaemia virus</td>
</tr>
<tr>
<td>NK</td>
<td>natural killer cells</td>
</tr>
<tr>
<td>NMDA</td>
<td>N-methyl-D-aspartate</td>
</tr>
<tr>
<td>NSI</td>
<td>non-syncytium inducing</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>OTUs</td>
<td>operational taxonimic units</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PBL</td>
<td>peripheral blood leukocyte</td>
</tr>
<tr>
<td>PBMC</td>
<td>peripheral blood mononuclear cell</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
</tbody>
</table>
PBS'  primer binding site
PCR  polymerase chain reaction
PHA  phytohaemagglutinin
PML  progressive multifocal leukoencephalopathy
RME  receptor mediated endocytosis
PR  protease
RNA  ribonucleic acid
RNA dep DNA pol  RNA dependent DNA polymerase
RNP  ribonucleo-protein complex
RRE  rev responsive element
RT  reverse transcriptase
SDS PAGE  sodium dodecyl sulphate PAGE
SI  syncytium inducing
SIV  simian immunodeficiency virus
SNV  spleen necrosis virus
ss  single stranded
SU  surface
TAR  trans-activation response element
TCID  tissue culture infective doses
TE  tris-EDTA buffer
TGF  transforming growth factor
TM  transmembrane
TNF  tumor necrosis factor
t-RNA  transfer RNA
<table>
<thead>
<tr>
<th>UPGMA</th>
<th>unweighted pair group method with arithmetic mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declaration</td>
<td>1</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>2</td>
</tr>
<tr>
<td>Abstract</td>
<td>3</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>6</td>
</tr>
<tr>
<td>Table of contents</td>
<td>11</td>
</tr>
<tr>
<td>List of Tables</td>
<td>15</td>
</tr>
<tr>
<td>List of Figures</td>
<td>16</td>
</tr>
<tr>
<td>Publications and presentation of work</td>
<td>18</td>
</tr>
<tr>
<td><strong>CHAPTER 1: INTRODUCTION</strong></td>
<td></td>
</tr>
<tr>
<td>1.1 Clinical aspects of HIV-1 infection.</td>
<td></td>
</tr>
<tr>
<td>1.1.1 Identification and isolation of HIV-1.</td>
<td></td>
</tr>
<tr>
<td>1.1.2 Classification of HIV-1 disease status.</td>
<td></td>
</tr>
<tr>
<td>1.1.3 Natural history of HIV-1 infection.</td>
<td></td>
</tr>
<tr>
<td>1.1.4 Monitoring the response to HIV-1 infection.</td>
<td></td>
</tr>
<tr>
<td>1.1.4.1 Diagnostic techniques.</td>
<td></td>
</tr>
<tr>
<td>1.1.4.2 Prognostic markers.</td>
<td></td>
</tr>
<tr>
<td>1.1.5 Transmission of HIV-1.</td>
<td></td>
</tr>
<tr>
<td>1.1.6 Epidemiology of HIV-1 infection.</td>
<td></td>
</tr>
<tr>
<td>1.2 Virological aspects of HIV-1 infection.</td>
<td></td>
</tr>
<tr>
<td>1.2.1 Classification of HIV-1.</td>
<td></td>
</tr>
<tr>
<td>1.2.2 Biochemical structure of HIV-1.</td>
<td></td>
</tr>
<tr>
<td>1.2.3 Genomic organisation of HIV-1.</td>
<td></td>
</tr>
<tr>
<td>1.2.4 Structural genes of HIV-1.</td>
<td></td>
</tr>
<tr>
<td>1.2.5 Regulatory genes of HIV-1.</td>
<td></td>
</tr>
<tr>
<td>1.2.6 Accessory genes of HIV-1.</td>
<td></td>
</tr>
<tr>
<td>1.2.7 Origin and evolution of HIV-1.</td>
<td></td>
</tr>
<tr>
<td>1.2.8 Life cycle of HIV-1.</td>
<td></td>
</tr>
<tr>
<td>1.2.8.1 Attachment.</td>
<td></td>
</tr>
<tr>
<td>1.2.8.2 Fusion and entry.</td>
<td></td>
</tr>
<tr>
<td>1.2.8.3 Reverse transcription.</td>
<td></td>
</tr>
</tbody>
</table>
1.2.8.4 Integration. 74
1.2.8.5 Synthesis of HIV-1 RNA. 75
1.2.8.6 Assembly and release. 76

1.2.9 Therapeutic strategies. 77
  1.2.9.1 Anti-retroviral agents. 77

1.3 Diversity of HIV-1. 81

1.4 Tropism of HIV-1. 89
  1.4.1 HIV-1 infection of different cell types. 89
  1.4.2 Differential properties in the host range of HIV-1. 93
  1.4.3 Role of the V3 region of gp120 in cellular tropism. 95
  1.4.4 Neurotropism of HIV-1. 99

1.5 Pathogenesis of HIV-1 infection. 101
  1.5.1 Mechanisms of HIV-1 induced cytopathicity. 102
    1.5.1.1 Direct mechanisms. 102
    1.5.1.2 Indirect mechanisms. 104
  1.5.2 Neuropathogenesis of HIV-1 infection. 110
  1.5.3 HIV-1 pathogenicity and disease progression. 113

1.6 Molecular evolution. 117
  1.6.1 Distance estimation. 117
  1.6.2 Rate of nucleotide substitution. 119
  1.6.3 Estimation of divergence times. 120
  1.6.4 Molecular phylogeny. 120
  1.6.5 Statistical significance of tree topologies. 123

AIMS. 125

CHAPTER 2: MATERIALS AND METHODS.

2.1 Clinical details of study patients. 127

2.2 Buffers and reagents.
  2.2.1 Extraction of DNA from tissue samples. 129
2.2.2 Amplification of proviral DNA.
2.2.3 Solid phase purification of PCR products.
2.2.4 Sequence analysis.

2.3 Extraction of DNA from tissue samples.
2.4 Quantitation of total extracted DNA.
2.5 Amplification of proviral DNA.
2.6 Limiting dilution assay.
2.7 Visualization of amplified PCR products.
2.8 Solid phase purification of PCR products.
2.9 Direct sequencing of amplified PCR products.
2.10 Length profiling analysis of amplified PCR products.
2.11 Nucleotide sequence accession numbers.

CHAPTER 3: INVESTIGATION OF THE DYNAMICS OF THE SPREAD OF HUMAN IMMUNODEFICIENCY VIRUS TO BRAIN AND OTHER TISSUES BY EVOLUTIONARY ANALYSIS OF SEQUENCES FROM THE p17 gag AND env GENES.

3.1 Introduction.
3.2 Clinical details of patient samples.
3.3 Phylogenetic analysis.
3.4 Statistical analysis.
3.5 Results.
3.5.1 Rate of sequence change in the p17 gag region.
3.5.2 Phylogenetic analysis of variants from different tissues.
3.5.3 Time of divergence of HIV-1 variants in different tissues.
3.5.4 Phylogenetic analysis of sequence variants from p79 isolated from different tissues.
3.6 Discussion.
3.6.1 Diversity of HIV-1 in vivo.
3.6.2 Rate of sequence change of HIV-1 in vivo.
3.6.3 Organ specific differences of HIV-1 in the V3 region.
3.6.4 Multiple evolutionary lineages in p17 gag and V1/V2 regions.

CHAPTER 4: INVESTIGATION OF POPULATION DIVERSITY OF HIV-1 IN VIVO BY NUCLEOTIDE SEQUENCING AND LENGTH POLYMORPHISM ANALYSIS OF THE V1/V2 HYPERVARIABLE REGION OF env.

4.1 Introduction.

4.2 Statistical analysis.

4.3 Results.
4.3.1 Intrasample variation in V1 and V2 hypervariable regions.
4.3.2 V1 and V2 sequence variability and tissue tropism.
4.3.3 Analysis of a number of V1/V2 length variants and disease status.

4.4 Discussion.
4.4.1 Lack of tissue specific grouping by sequence variants from V1 and V2 hypervariable domains.
4.4.2 Relationship between V1 and V2 sequences with tissue distribution.
4.4.3 Does the diversity of V1 and V2 sequence variants correlate with disease progression?

CHAPTER 5: GENERAL DISCUSSION.

BIBLIOGRAPHY.

APPENDICES.
LIST OF TABLES

Table 1: Classification system for HIV infection. 23

Table 2: Classification of retroviruses. 45

Table 3: Standardized nomenclature for HIV proteins. 49

Table 4: Potential targets for therapeutic drugs. 78

Table 5: Clinical details of study patients. 128

Table 6: Sequence comparisons between study subjects in the pl7gag region. 153

Table 7: Sequence comparisons of variants from different tissues in the pl7gag region. 159

Table 8: Sequence comparisons of variants from different tissues from p79 in the pl7gag region. 165

Table 9: Comparisons of V1 and V2 obtained by LPA compared with direct sequencing in the V1 (A) and V2 (B) regions. 202
LIST OF FIGURES

Figure 1: Clinical course of HIV-1 infection 24F

Figure 2: Biochemical structure of HIV-1. 46F

Figure 3: Genomic organization of HIV-1. 50F

Figure 4: Secondary structural model of gp120. 53F

Figure 5: Life cycle of HIV-1. 64F

Figure 6: Clinical and laboratory markers of disease progression in the four study patients in the 5 years before death. 148F

Figure 7: Neighbor-joining tree of sequences from the p17\textsubscript{gag} region of three of the study subjects (p4, p5 and p6). 151F

Figure 8: Frequency histograms of silent pairwise distances in the p17\textsubscript{gag} region between study subjects p4 and p5 (A), p4 and p6 (B), and p5 and p6 (C) and within study subjects between sequences obtained from different tissues of p4 (D), p5 (E) and p6 (F). 153F

Figure 9: Phylogenetic analysis of sequences obtained from different tissues of three of the study subjects (p4 [A and D], p5 [B and E], and p6 [C and F]) in different subgenomic regions (p17\textsubscript{gag} [A through C] and V1/V2 flanking regions [D through F]). 155F

Figure 10: Distribution of pairwise distances in different tissues from three of the study patients (p4, p5 and p6) at silent (A) and nonsilent (B) sites in the p17\textsubscript{gag} region. 159F

Figure 11: Phylogenetic analysis of sequences obtained from different tissues of p79 in the p17\textsubscript{gag} region. 162F

Figure 12: Distribution of pairwise distances in different tissues from p79 at silent (A) and nonsilent (B) sites in the p17\textsubscript{gag} region. 165F

Figure 13: Proviral V3 loop amino acid sequences from infected organs from terminal AIDS patients (p4-6, 9). 174F

Figure 14: Comparison of the relationship between charge, length and glycosylation of V1 and V2 domains with the experimentally determined in vitro phenotype of previously described HIV isolates. (A), (B): charge and length for V1 and V2 domain respectively; (C), (D): number of potential N-linked glycosylation sites for V1 and V2 domains respectively. 187F
Figure 15: Proviral V1 and V2 domain amino acid sequences from three study subjects (A) p4; (B) p5; (C) p6.

Figure 16: Comparison of overall charge of V1 and V2 sequence variants from p4 (A, D), p5 (B, E), p6 (C, F) in V1 (A, B, C) and V2 (D, E, F).

Figure 17: Comparison of overall length of V1 and V2 sequences from p4 (A, D), p5 (B, E), p6 (C, F) in V1 (A, B, C) and V2 (D, E, F).

Figure 18: Comparison of the number of potential N-linked glycosylation sites in V1 and V2 sequence variants from p4 (A, D), p5 (B, E), p6 (C, F) in V1 (A, B, C), and V2 (D, E, F).

Figure 19: Neighbour-joining trees of sequences in V1 and V2 domains of three study subjects (A) p4; (B) p5; (C) p6.

Figure 20: LPA of HIV variants obtained from lymph node and brain tissue of p4, p5 and p6 for (A) V1 and (B) V2 regions.

Figure 21: Plot of CD4 lymphocyte count -v- population diversity as determined by LPA in V1 (A) and V2 (B) regions. Comparisons of the distribution of length variants observed in different tissues in V1 (C) and V2 (D) regions.
PUBLICATIONS AND PRESENTATION OF WORK.

PUBLICATIONS.

Hughes, E. S., J. E. Bell, and P. Simmonds. (1997). Relationship between sequence diversity in the V1/V2 hypervariable region of the human immunodeficiency virus type 1 (HIV-1) env gene with the distribution of HIV-1 in different tissues in vivo (Journal of General Virology, in press).


ORAL PRESENTATIONS.

Society for General Microbiology, University of Warwick, March 1996. Title: Investigation of the dynamics of HIV spread to the brain by phylogenetic analysis of sequences in p17_gag and V1 and V2 regions from env.

POSTER PRESENTATIONS.


MRC- 8th Annual AIDS Workshop, The Manchester Conference Centre, September 1994. Title: Investigation of in vivo sequence variation of V1 and V2 hypervariable regions within lymphoid and non-lymphoid tissues upon disease progression.
CHAPTER 1. INTRODUCTION: HUMAN IMMUNODEFICIENCY TYPE 1 (HIV-1).
1.1 CLINICAL ASPECTS OF HIV-1 INFECTION.

1.1.1 IDENTIFICATION AND ISOLATION OF HIV-1.

Infection with the human immunodeficiency virus type-1 (HIV-1) is associated with a slow, progressive and irreversible impairment of the immune system eventually leading to severe immunodeficiency clinically known as AIDS. It is now universally accepted that HIV-1 is the main aetiological agent of the acquired immune deficiency syndrome (AIDS). Studies to determine how widely disseminated the virus was eventually led to the discovery of a second related virus, HIV-2, which was found to be capable of inducing the same clinical symptoms in humans as HIV-1 (Clavel et al., 1986; Barin et al., 1985).

Approximately 15 years ago, Gottlieb and colleagues reported an unusually high number of young homosexual males presenting with Pneumocystis carinii pneumonia and other unusual opportunistic infections (Gottlieb et al., 1981). Following this report a succession of observations involving young male homosexuals presenting with unusual opportunistic infections and immune deficiencies were reported (Stahl et al., 1982; Friedman Kien et al., 1982; Klein et al., 1984; Metroka et al., 1983; Mildvan et al., 1982; Siegal et al., 1981; Drew et al., 1981; Masur et al., 1981). Thus far, this new illness was considered by most to be a 'gay' disease. In 1981 this issue was further confused when an increasing number of Haitian immigrants to the United States were reported to have died with
toxoplasmosis and cytomegalovirus in their brains (Moskowitz et al., 1983) and was further compounded by the appearance of cases in women. The juxtaposition of these various reports proved to be fortuitous as events following these observations led to the conclusion that a single infectious agent was responsible for this epidemic of immune deficiency.

Isolation of this infectious agent was finally accomplished in 1983 when the virus was cultured in T lymphocytes, from a lymph node biopsy of a homosexual patient with multiple lymphadenopathies (Barre-Sinoussi et al., 1983). It was suggested that this novel virus was a member of the human T-cell leukaemia virus (HTLV) family. Serum from this patient strongly reacted with surface antigens of HTLV-1 and the divalent cationic requirements of the viral reverse transcriptase (RT) were similar to that for HTLVs. However, it was thought to be a distinct virus since core proteins of this novel virus were immunologically unrelated to those of HTLV-1 (Barre-Sinoussi et al., 1983). This virus was further identified as a retrovirus using a sucrose density gradient and electron microscopy illustrating particles budding from the surface of umbilical cord lymphocytes characteristic of retroviruses. This virus was therefore designated lymphadenopathy-associated virus (LAV). An association of members of the HTLV family with patients with AIDS was subsequently reported from numerous independent sources (Levy et al., 1984; Essex et al., 1983; Gelman et al., 1983; Gallo et al., 1984). For some time there was a certain degree of confusion regarding the nomenclature of this virus implicated in AIDS. Initially there were three claims on the nomenclature: LAV (Barre-Sinoussi et al., 1983), HTLV-III (Gallo et al., 1984) and AIDS related virus
(ARV; Levy et al., 1984). However, it is now universally known as the human immunodeficiency virus type-1 (HIV-1; Coffin et al., 1986b).

1.1.2 CLASSIFICATION OF HIV-1 DISEASE STATUS.

Infection with HIV-1 can result in a spectrum of clinical conditions, ranging from asymptomatic infection to severe immunodeficiency and the acquisition of numerous opportunistic infections and neoplasms. The Centres for Disease Control (CDC) have constructed a classification system which catalogues the manifestations of HIV-1 into four clinical stages: Group I-IV (CDC, 1986; Table 1). CDC group I identifies those patients who have recently undergone seroconversion, shortly following initial infection, which may be accompanied by an acute seroconversion illness, with clinical features such as a mononucleosis-like illness, fever and malaise (Cooper et al., 1985). CDC group II includes asymptomatic patients showing no signs or symptoms of HIV-1 infection. This asymptomatic period can last from a few months to a number of years before any signs of clinical disease are apparent. Patients presenting with persistent generalized lymphadenopathy are classified as CDC group III. CDC group IV describes a number of clinical manifestations and has been subdivided into five subgroups A to E. Subgroup A (constitutional disease) includes individuals with one or more of the following afflictions: fever (> 1 month), diarrhoea (> 1 month) and/or involuntary weight loss (> 10% of baseline). Infection must be unattributable to any infection other than HIV infection. Subgroup B (neurological disease) includes individuals suffering
### TABLE 1: CLASSIFICATION SYSTEM FOR HIV INFECTION

<table>
<thead>
<tr>
<th>CDC group</th>
<th>Clinical manifestation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td>Acute infection</td>
</tr>
<tr>
<td>Group II</td>
<td>Asymptomatic infection</td>
</tr>
<tr>
<td>Group III</td>
<td>Persistent generalized lymphadenopathy</td>
</tr>
<tr>
<td>Group IV</td>
<td>Other diseases</td>
</tr>
<tr>
<td>Subgroup A</td>
<td>Constitutional disease</td>
</tr>
<tr>
<td>Subgroup B</td>
<td>Neurological disease</td>
</tr>
<tr>
<td>Subgroup C</td>
<td>Secondary infectious diseases</td>
</tr>
<tr>
<td>Category C-1</td>
<td>Specified secondary infectious diseases listed in the CDC surveillance definition for AIDS&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Category C-2</td>
<td>Other specified secondary infectious diseases&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Subgroup D</td>
<td>Secondary cancers&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Subgroup E</td>
<td>Other conditions</td>
</tr>
</tbody>
</table>

<sup>a</sup>Appendix I  
<sup>b</sup>Appendix II  
<sup>c</sup>Appendix III
from one or more of the following neurological diseases: dementia, myelopathy or peripheral neuropathy. Infection must solely be due to invasion of the nervous system by HIV and not attributable to any other illness or condition. Subgroup C is divided into two categories (C1 and C2) and includes individuals with clinical disease associated with secondary infectious diseases and a defect in cell-mediated immunity. Subgroup D is associated with the development of secondary cancers and a defect in cell mediated immunity. Subgroup E includes other clinical disorders, previously unclassified, which may be attributable to HIV infection. The CDC produced a revised classification system that extended the diagnosis of AIDS to asymptomatic individuals with CD4+ T-cell counts of < 200 cells/µl. Similarly, the number of previously defined clinical categories, group IV, has been expanded to include further clinical conditions (CDC. 1992).

1.1.3 NATURAL HISTORY OF HIV-1 INFECTION.

The natural history of HIV-1 infection can be viewed as a progression of three distinct clinical stages: an initial acute stage associated with primary infection (CDC stage I), a chronic stage resulting in a period of clinical, but not virological, latency (CDC stages II and III) and finally, a crisis stage representing a period of profound immunodeficiency manifest by the development of opportunistic infections and neurological disorders (CDC stage IV; Fig 1).

During primary infection, plasma HIV-1 titres increase dramatically manifested by high levels of viraemia and serum p24 antigen (Simmonds, 1990a;
Fig. 1. Clinical course of HIV-1 infection (modified from Medical Microbiology, 2nd edition, 1996).
Acute infection

Chronic lymphadenopathy

Latency

Subclinical immune dysfunction

Systemic immune deficiency

Mucous membrane defects

AIDS

AIDS dementia

Anti-HIV-1 antibody

CD4+ T cell count

Virus

Time (months)

CD4+ T cell count (cells/mm³)

CDC stage I

CDC stage II

CDC stage III

CDC stage IV
This plasma viraemia is only transient (up to $10^3$ to $10^4$ tissue culture infective doses (TCID) per ml have been detected in plasma; Clark et al., 1991; Daar et al., 1991), and generally only lasts for a few weeks subsiding when a specific antibody response is mounted. However, a cellular cytotoxic immune response almost certainly contributes to the initial containment of primary HIV-1 infection. The immune mechanisms responsible for this viral clearance are discussed in detail below. In addition, the acute stage of infection may coincide with an acute seroconversion illness characterized by fever, rash and myalgia (Clark et al., 1991; Daar et al., 1991). CD4 cell counts fall dramatically during this stage of infection, although it is not clear whether this is a result of a cytopathic effect by HIV-1 or redistribution to extravascular sites (Fahey et al., 1990; Bofill et al., 1996; Ferbas et al., 1996). Following the curtailment of acute primary infection a chronic, clinically asymptomatic stage ensues. This is often termed the latent period and although clinically latent recent studies have revealed HIV-1 infection as a dynamic process in which a stable steady state hides a very high rate of viral replication and turnover of infected cells (Wei et al., 1995; Ho et al., 1995; Ho et al., 1989; Coombs et al., 1989; Saag et al., 1991; Michael et al., 1992). Similarly, a number of studies have reported active replication of HIV-1 in lymphoid tissue (Pantaleo et al., 1991; Graziosi et al., 1993; Pantaleo et al., 1993). During this chronic stage of persistent infection only low levels of plasma viraemia can be detected ($30$ TCID$_{50}$ per ml; Ho et al., 1989) and few infected CD4+ cells are present in the peripheral
blood (1 in 10,000; Harper et al., 1986). As disease progresses the development of clinical immunodeficiency follows the reappearance of significant levels of plasma viraemia and an increase in the expression of HIV-1 DNA and RNA (Michael et al., 1992), which is associated with an accelerated depletion of CD4+ cells (MacDonell et al., 1990; Ho et al., 1989; Coombs et al., 1989; Saag et al., 1991; Ferbas et al., 1996). As the number of CD4+ cells decreases so the percentage of infected lymphocytes increases. Indeed, in one study of patients with clinical AIDS the number of infected lymphocytes was found to exceed 1 in 40 (Ho et al., 1989). Similarly, a number of studies have shown that the number of CD4+ cells decreases as the plasma viraemia increases and may exceed $10^4$ TCID$_{50}$ per ml (Coombs et al., 1989; Ho et al., 1989; Saag et al., 1991). Saag et al., also studied the clinical course of infection in children infected perinatally with HIV-1 and found that they frequently exhibited significantly high plasma levels and accelerated disease irrespective of CD4+ cell counts (Coombs et al., 1989). This is most likely due to the immature nature of these childrens’ immune response failing to control the initial viraemia.

The rapid decline in viraemia after primary infection suggests a potent antiviral role for the host immunity. A number of studies suggest that HIV-1 specific cytotoxic T lymphocytes (CTL) may play an important role in the clearance of acute infection (Rouse et al., 1988; Plata et al., 1987; Walker et al., 1987). Following HIV-1 infection CTL precursors (CTLp) are produced in regional lymph nodes followed by the accumulation of virus specific CD8+ CTL at sites of virus replication. This CTL activity reaches a peak shortly after viral replication,
and in some cases prior to a specific antibody response (Koup et al., 1994; Walker et al., 1986; Walker et al., 1990). Koup et al., detected CTLp specific for cells expressing HIV-1 gag, pol and env antigens within three weeks of infection and found that they remained elevated for at least three to six months following acute viral infection (Koup et al., 1994). The CTL response is the first virus specific immune response detected suggesting CTL activity is responsible for viral clearance and predict the consequences of the acute stage of infection. Neutralizing antibodies are undetectable until several weeks after viraemia has subsided. It is possible that CTL are not the only mechanism involved in the clearance of acute viraemia during primary HIV-1 infection. Natural killer (NK) cells are also likely to play a role although they have not been studied in detail. However, they are important in the clearance of other acute viral infections such as acute infectious mononucleosis (Tomkinson et al., 1989; Nahill et al., 1993), and transient elevations in their numbers have been observed in acute HIV-1 infection (Koup et al., 1994).

Following seroconversion, approximately four weeks after infection, antibodies to different structural components of HIV-1 can be detected by enzyme immunoassays (EIA). These do not appear simultaneously and antibodies to the envelope glycoprotein, gp120, are the first to be detected. Subsequently, antibodies to the regulatory proteins, p24 from gag and p53 and p64 from pol, are produced. Antibodies to the transmembrane glycoprotein, gp41, are often not detected until several weeks later. Although detection of antibodies to HIV-1 structural and regulatory proteins are indicative of infection, their role as an antiviral component
of the immune system is less certain. The presence of anti-HIV-1 neutralizing antibodies in infected individuals has been reported by several investigators. The envelope glycoproteins gp120 and gp41 are among the principal targets for the host humoral immune response to HIV-1 (Fung et al., 1992; Matsushita et al., 1988; Javaherian et al., 1990; Ho et al., 1991; Looney et al., 1988; Haigwood et al., 1990; Brolden et al., 1991; Weiss et al., 1986). Virtually all neutralizing activity in the sera of infected individuals is directed against these proteins and particularly against regions of gp120. The V3 hypervariable region has been found to be the principal neutralization determinant in HIV-1 infection (Kuiken et al., 1993; Wolfs et al., 1991; Zwart et al., 1991; Javaherian et al., 1990; Javaherian et al., 1989; Skinner et al., 1988). V3 neutralizing antibodies can prevent HIV-1 entering target cells, although they do not necessarily abrogate binding to the cell via the gp120-CD4 interaction (Linsley et al., 1988; Skinner et al., 1988). The V3 loop can present both linear and conformational determinants for antibody recognition. A second major neutralizing region is the CD4 binding site located in the carboxy terminal region of gp120 (Pinter et al., 1993; Ditzel et al., 1995; Kang et al., 1991). This determinant is generally conformation-dependent and antibodies block the binding of HIV-1 to the cellular receptor. Neutralizing antibodies directed against the V1 and V2 hypervariable regions of gp120 have also been detected (Mankowski et al., 1994; Warrier et al., 1994; McKeating et al., 1993; Gorny et al., 1994; McKeating et al., 1993; Moore et al., 1993). As with the V3 loop, linear and conformational determinants appear to be involved. The envelope glycoprotein gp41 has not been analysed as extensively as gp120 although, neutralizing
antibodies have been detected (Purtscher et al., 1994; Sattentau et al., 1995; Purtscher et al., 1996; Back et al., 1993). Generally neutralizing antibodies to gp41 are recognized following the gp120-CD4 interaction suggesting that receptor binding induces conformational changes exposing the gp41 domain (Sattentau et al., 1993). The role of neutralizing antibodies in clearing viraemia during the acute phase of infection remains controversial. Several studies have found the decline in viraemia precedes the production of neutralizing antibodies, although it is possible that antibodies are present at very low levels prior to the decline in viraemia undetectable by current assays.

1.1.4 MONITORING THE RESPONSE TO HIV-1 INFECTION.

1.1.4.1 DIAGNOSTIC TECHNIQUES.

The diagnosis of infection with HIV-1 is dependent on the detection of specific HIV antibodies, detection of circulating viral antigens, isolation of the virus from clinical specimens, or detection of viral genetic material. Routine diagnostic procedures used to identify HIV-1 infected individuals rely largely on the presence of antibodies to viral antigens in the serum. Antibody tests combining a screening assay and a confirmatory test have been the most widely used techniques for establishing the presence of HIV-1 infection (Phair et al., 1992). Most HIV antibody screening assays use the rapid and convenient EIA. Originally EIAs used whole viral lysates obtained from cultivation of HIV in lymphocytes
(Jackson et al., 1988; Johnson, 1992). These preparations however, resulted in false positive results due to the reaction of serum antibodies to cellular components. Subsequent assays have been developed that use recombinant HIV-1 proteins to env and gag products and synthesized peptides corresponding to conserved regions of env and p24 core proteins (Johnson, 1992). These assays have proven to be more sensitive and specific than the original EIAs.

Diagnosis of HIV-1 infection is not always possible based on the presence of antibodies in the serum. For example antibodies are not present in the acute phase of infection (serological window period) and can be lost in the advanced stages of disease. Therefore diagnostic tests for either direct detection of the virus or its components may be important in monitoring the disease when no antibody is present. Virus culture can provide definitive information about the presence or absence of HIV-1 infection. However it is costly, time consuming and labour intensive with the potential for exposure to high concentrations of infectious virus. Therefore although it is a good reference method for identifying HIV-1 infection it is of limited practical use for diagnosis, and is not suitable for measurement of circulating virus load.

HIV-1 antigenaemia is thought to be significant during two periods of HIV-1 infection, initially during the acute phase, prior to the production of antibodies, and at the terminal stages of infection and development of AIDS when virus production increases dramatically (Gaines et al., 1987; von Sydow et al., 1988; Pedersen et al., 1987). During these brief periods p24 antigen can be detected in both serum and plasma using an EIA based antigen capture technique (Goudsmit et
However, direct detection of p24 antigen in infected sera is unreliable following seroconversion in most cases because of the small amount of circulating antigen (MacDonell et al., 1990). In general p24 antigen cannot be detected in the serum of 85% or more asymptomatic HIV-1 seropositive individuals (Fahey et al., 1990). The low detection rate of p24 antigen in serum from asymptomatic patients may be due to low levels of antigen production or to the formation of immune complexes. In order to increase the sensitivity of this diagnostic test Nishanian et al., developed an acid dissociation procedure designed to disrupt p24 antigen-antibody immune complexes (Nishanian et al., 1990). This method improved the detection and quantitation of p24 antigen EIA testing allowing a more accurate assessment of in vivo viral replication.

In more recent years, PCR-based methods have been developed to quantify specific DNA and RNA sequences present in small amounts of biological specimens. These methods were developed as an alternative to conventional diagnostic molecular hybridization methods such as Southern and Northern blots, dot hybridization and spot hybridization, as these methods show inadequate sensitivity for HIV in vivo. PCR based applications provide a specific and efficient method for direct detection of HIV-1 infection and can detect DNA and RNA in plasma of HIV-1 infected individuals regardless of clinical condition (Zhang et al., 1991; Piatak et al., 1993; Ottmann et al., 1991). This technique will therefore be valuable for identifying HIV-1 infected individuals prior to seroconversion and identifying infected babies born to HIV-1 infected mothers in whom maternal antibodies can persist for 5 to 14 months (Rogers et al., 1989). Similarly, this
technique has important advantages for resolving equivocal serologic results from HIV-1 infected patients (Jackson et al., 1990).

1.1.4.2 PROGNOSTIC MARKERS.

Numerous markers have been studied in patients infected with HIV-1 in an effort to monitor and predict progression to disease. Monitoring of HIV-1 infection can be classified into three categories: the direct detection of the virus or its antigens, methods which quantitate the immunosuppression of HIV-1 and evaluation of the immune response to HIV-1.

Previously detection of infectious HIV-1 in plasma was carried out following co-cultivation with phytohaemagglutinin (PHA) stimulated donor PBMCs (Coombs et al., 1989). Quantification of plasma viraemia was carried out using serial dilution co-culture and detection of p24 antigen production \textit{in vivo} (Ho et al., 1989; Coombs et al., 1989). These studies concluded that plasma viraemia correlated well with progression to disease since plasma viraemia was detected more frequently in symptomatic individuals than in asymptomatic individuals.

More recently a more sensitive quantitative approach has been developed using a novel PCR based method. Several investigators have coupled a reverse transcriptase (RT) reaction step to conventional RNA PCR and have detected HIV RNA in HIV-1 cell cultures and PBMCs from infected individuals (Arrigo et al., 1989; Byrne et al., 1988). These studies have confirmed that a low level of viral replication occurs throughout the clinically latent stage of infection, although they do not determine
absolute quantities of RNA. In order to quantify actively replicating virus a number of methods have been proposed including competitive RT-PCR (cRT-PCR) and a limiting dilution assay (Zhang et al., 1991; Menzo et al., 1992). Both studies provide evidence that quantitation of plasma viraemia using an RT-PCR approach is a more accurate method for the detection and quantitation of virus production. Therefore, these methods provide a better marker for the progression of disease, in contrast with p24 antigen assays which are limited by the production of immune complexes preventing the detection of p24 antigen.

During HIV-1 infection and progression to AIDS a marked decrease in CD4+ T cells is apparent. These cells are central to the regulation of the immune system and the sharp decline in numbers is thought to be central to the resulting immunodeficiency. During the asymptomatic phase there appears to be an increase in CD8+ T cells (cytotoxic) suggesting that this cell subset may contribute to the limiting of HIV-1 replication during this phase (Giorgi et al., 1987; Lang et al., 1989). An initial inversion of the CD4+:CD8+ T cell ratio is another prominent feature of HIV-1 infection, although it is unclear whether this inversion is due to a decrease in the number of CD4+ T cells or augmentation of CD8+ T cells. Many studies have been carried out to determine the prognostic value of these immunological markers (Miedema et al., 1990a; Giorgi et al., 1987a; Fahey et al., 1984a; Bogner et al., 1991a; Lang et al., 1989a; Phillips et al., 1991a). Results from these studies indicate that following seroconversion a rapid decrease in CD4+ T cells occurs as a result of the pathogenic effects of HIV-1. This is followed by a continual fall in CD4+ T cells but at a much slower rate and for a variable period
of time during the asymptomatic phase. Prior to the progression to AIDS (approximately two years) CD4+ T cell counts begin to fall more rapidly again and low counts are common just prior to progression to AIDS. There are various methods available to monitor the CD4+ T cell count: the absolute CD4+ T cell number in blood, the percentage of CD4+ T cells and the CD4+:CD8+ T cell ratio. It is now widely accepted that CD4+ T cell counts are a reliable prognostic marker to determine progression to AIDS and may be a suitable marker for monitoring the effects of antiviral treatments.

There are various soluble products of immune activation present in the serum that can be readily measured. Beta-2-microglobulin (β-2-M) is a low molecular weight protein (11,800 Da) that forms the light chain of the class I major histocompatibility complex found on the surface of most nucleated cells (Jacobson et al., 1989; Hofmann et al., 1990). β-2-M is normally present in the serum and urine and normal levels reflect the balance between production and removal of this molecule. Serum levels are generally increased in diseases characterized by increased cell turnover including malignancies and chronic inflammation. It is also associated with renal failure and has been reported to be elevated in various diseases affecting lymphocytes, such as cytomegalovirus, hepatitis and HIV infection (Hofmann et al., 1990; Fuchs et al., 1991). Due to the immunosuppression caused as a result of HIV-1 infection and a reduction in CD4+ lymphocytes β-2-M was considered to be a good prognostic marker during progression to AIDS. Various investigations have been carried out to determine how infection with HIV-1 affects the levels of β-2-M (Mastroianni et al., 1990;
Jacobson et al., 1989; Hofmann et al., 1990; Fuchs et al., 1991). Analysis of asymptomatic and symptomatic individuals has shown that β-2-M levels increase during progression to AIDS. Similarly, analysis of HIV-1 seroconverters showed that changes in β-2-M levels after exposure to HIV-1 were associated with the decline in CD4+ T cells.

Neopterin is another serum marker associated with immune cell activation, a low molecular weight compound derived biosynthetically from guanosinetriphosphate (Fuchs et al., 1988b). It is predominantly produced following macrophage stimulation with interferon gamma (IFN-γ) released from activated T cells (Huber et al., 1984). Increased neopterin levels have been demonstrated in individuals with diseases associated with or caused by cell mediated immunity (CMI), including allograft rejections, autoimmune disorders, viral infections, bacterial infections, parasitic infections and various malignancies (Fuchs et al., 1991; Fuchs et al., 1988). Neopterin levels behave similarly to β-2-M during the course of HIV-1 infection and parallel disease progression. Various studies have investigated the levels of neopterin in asymptomatic and symptomatic individuals and found much higher levels of this molecule in symptomatic individuals (Fuchs et al., 1991; Fuchs et al., 1988a; Fuchs et al., 1988b). Detection of neopterin was found to be as sensitive a prognostic marker as the CD4+ T cell count and increased levels of this molecule correlated with the rapid decrease in CD4+ T cell counts. Both β-2-M and neopterin have been found to have good predictive power, probably on a parallel with CD4+ T cell counts, acting independently of each other and indeed of CD4+ T cell measurements. Therefore the level of β-2-M or
neopterin used in conjunction with CD4+ T cell counts may provide a better method for monitoring progression to disease than CD4+ T cell counts alone.

1.1.5 TRANSMISSION OF HIV-1.

HIV transmission can primarily be defined into three major groups depending on the mode of transmission: (1) unprotected sexual intercourse, (2) inoculation of infected blood or blood products and, (3) vertically from mother to child. The risk factors involved in transmission of HIV in each of these three groups vary depending on the nature of the activity through which transmission occurs. However, in each case one unifying risk factor exists, that is the stage of disease. An HIV infected individual appears to be more infectious in the very early stages (prior to antibody production) and again in the late stages of clinical AIDS due to higher levels of virus present in the blood during these stages of disease than during the asymptomatic stage (European study group., 1989; Jacques et al., 1994; Lazzarin et al., 1991).

Transmission of HIV through unprotected sexual activity can occur from male to male, male to female and female to male (Caceres et al., 1994; European Study Group, 1992). Female to female transmission remains extremely rare (Raiteri et al., 1994). A number of factors influence transmission of HIV through sexual contact. Generally the receptive partner in either homosexual or heterosexual activity is more susceptible to HIV infection (Rodrigues et al., 1995). Anal intercourse carries the risk of trauma to the mucosal lining and hence increased risk
of infection. During heterosexual intercourse male to female transmission is more effective than female to male transmission (European study group., 1989; European study group., 1992). The frequency of sexual contacts is thought to be the most important factor in creating the worldwide epidemic of sexually transmitted HIV infection, although a single contact can be sufficient to transmit HIV. An important biological factor increasing the probability of HIV transmission is the presence of variable sexually transmitted diseases, such as genital ulcers (Rodrigues et al., 1995).

The second mode of transmission is through blood and blood products. Contaminated blood is highly infectious upon parenteral exposure. The primary risk group for HIV transmission via blood is through intravenous drug abuse (Des Jarlais et al., 1988). This route of transmission is significant due to an epidemic of drug abuse and a change from oral to intravenous administration over the past few decades. Also an increase in needle sharing behaviour, especially in countries where 'shooting galleries' exist, provides a fertile field for the transmission of HIV among a large network of drug injectors. Also exposure is often repeated due to the addictive nature of these drugs. There is also an increased risk of transmission to sexual partners of intravenous drug users (Ronald et al., 1993). Similarly, frequent drug abuse is associated with other high risk behaviours such as sexual promiscuity. For example, prostitution is frequently motivated by the cost of drug addiction (Celentano et al., 1994).

Transmission of HIV-1 to blood transfusion recipients and haemophiliacs receiving blood clotting factors, such as factor VIII and IX, was originally a
significant risk of infection. However, this route of transmission is now very rare in countries where blood is screened for HIV antibodies and viral inactivating steps, such as heat inactivation, are routinely carried out in blood product manufacture, such as factors VIII and IX (Lackritz et al., 1995; Mannucci, 1993). Unfortunately contaminated blood is still a route of infection for HIV in some developing countries where difficulties have been encountered in setting up safe blood transfusion systems (Lackritz et al., 1993). Health care workers have been infected after a needle stick injury from a needle containing HIV infected blood or, less frequently, through an open cut or splashes into a mucous membrane (e.g. eyes, inside of nose). To date there has only been one case where a health care worker has infected patients. This involved the transmission of HIV from a dentist to a number of patients (Ou et al., 1992). Other practices which hold a potential risk of transmission through blood include those that involve piercing of the skin which acts as a natural barrier to infection. For example, infection may occur through tattooing, acupuncture or the use of other invasive instruments that may have come in contact with contaminated blood and have not been properly disinfected.

The third mode of transmission is vertically from an infected mother to her child. The parameters involved in this mode of transmission are not well defined. It is thought infection may occur in utero (intrauterine or transplacental; Conaldi et al., 1995), during delivery through the birth canal (perinatal) or after birth through breast milk (infancy; Palasanthiran et al., 1993; Looney et al., 1988; Toniolo et al., 1995). Transmission from mother to child ranges from 15% to 35% with the lowest rates reported in Europe and the highest in Africa (Stlouis et al., 1993; Newell et
1.1.6 EPIDEMIOLOGY OF HIV-1 INFECTION.

The global pandemic of HIV infection comprises many separate epidemics in different geographical areas and is influenced by many factors including the time of introduction of the virus, population density and diverse cultural and social variables. Previously HIV infection in distinct geographical areas could be separated according to the major mode of transmission. However this distinction is becoming less relevant as the epidemic develops. For example, in Europe the main mode of transmission was initially through homosexual contacts. Increasingly, however transmission through drug abuse and heterosexual routes are becoming apparent. Similarly, in South America the initial AIDS epidemic was focused in homosexual communities but heterosexual transmission is now considered to be a major route of exposure to HIV infection. Spread of HIV infection has varied considerably between developed and developing countries, depending on cultural, social and behavioural patterns.

In Africa, HIV infection is the leading cause of adult death in many cities and has greatly increased childhood mortality (Gregson et al., 1994). The World Health Organization (WHO) has estimated that approximately 16 million individuals in Africa have been infected with HIV, most of them in sub-Saharan Africa (World Health Organization, 1995). Heterosexual transmission is the main mode of transmission among African adults. As a result of the increasing numbers
of women becoming infected through heterosexual transmission, the prevalence of paediatric infections from mother to child is extremely high and expected to rise.

In most Asian countries, HIV infection was not recognised until much later than in the USA or Africa. However over 4 million people are now thought to be infected (Quinn, 1996). Thailand was the first Asian country in which HIV infection was recorded. Transmission was notably through drug abuse and high risk heterosexual activity, such as female sex workers (Morris et al., 1996). HIV seroprevalence increased from 1.2% to 45% in intravenous drug users (IDU) between 1988 and 1991 in Thailand (Quinn, 1996). In addition, HIV prevalences of 43% to 82% were detected in IDU in China (Cheng et al., 1994). HIV seroprevalence among female sex workers and their contacts are also surprisingly high with a prevalence of 30% to 65% reported among female sex workers in Thailand and India (Bollinger et al., 1995; Kaldor et al., 1994). A similar pattern to that observed in Thailand and India may be expected in many other countries in Asia that share the same prevalence of female sex workers and high frequency of IDU.

The number of infected individuals in the Pacific Islands has been estimated to be about 75 000, with most in Australia and New Zealand (World Health Organization. 1995). Most infections in these regions have occurred among homosexual men. The frequency of infection among IDU is relatively low compared to Europe and North America although this mode of transmission remains a significant route for the dissemination of HIV infection.

HIV infection amongst individuals in the Americas has been estimated at
three million, with over one million in North America and two million in Latin America and the Caribbean (World Health Organization. 1995; Quinn, 1996). There is a trend in these countries for a marked decline of HIV-1 among homosexual individuals and a substantial increase in the heterosexual population (Heverkos et al., 1995). For individual countries, the highest number of reported cases in the world originate in the USA with over 500 000 cases by the end of 1995 (Centres for Disease Control and Prevention. 1995).

Half a million individuals are thought to be infected in Western and Eastern Europe with transmission routes differing significantly between countries. For example, in Italy, Spain and Poland almost two thirds of infections are attributable to IDU, whereas homosexual individuals constitute the majority of infections in Scandinavia, Czech and Slovak republics. In Europe as a whole, the number of cases attributable to homosexual transmission decreased significantly from 62% to 36% between 1985 and 1992 (Quinn, 1996). The opposite trend for IDU has been reported with an increase from 16% to 40% in recent years. Similarly, transmission through heterosexual intercourse has shown an increase and seroprevalence studies indicate that between 14% and 18% of infected individuals may have acquired HIV infection heterosexually.

Future predictions of the global pandemic of HIV infection are very difficult to make with any degree of confidence. However, predictions by the WHO are that there will be approximately 26 million HIV infected individuals by the year 2000 (Quinn, 1996). It is suspected that the incidence of infection in younger populations (15 to 25 years old) will increase, as is already evident in the USA and
sub-Saharan Africa. The number of cases in sub-Saharan Africa and Asia are expected to continue to rise although, in America and Europe the number of HIV infected individuals are expected to remain fairly constant as the number of new infections are likely to parallel the number of fatalities. Therefore, future efforts to control the AIDS epidemic must be focused on educating people, to influence change in social and cultural behaviour until an effective vaccine regime or better therapeutic interventions are available.

1.2 VIROLOGICAL ASPECTS OF HIV-1 INFECTION.

1.2.1 CLASSIFICATION OF HIV-1.

HIV is a member of a large group of viruses known as the Retroviridae. Retroviruses display a variety of common features specific to this virus group. They are surrounded by a spherical envelope, which is acquired as the virion buds from the plasma membrane. The envelope surrounds a core protein encapsidating two copies of a positive strand ribonucleic acid (RNA) genome. The genome of retroviruses possess a common viral structure organised into three polyprotein genes: group specific antigens (gag), polymerase (pol) and envelope (env). The life cycle of a retrovirus involves the insertion of the viral genome into the host genetic material utilizing a virally encoded enzyme, integrase (IN). Replication proceeds through a deoxyribonucleic acid (DNA) intermediate again utilizing a virally encoded RNA-dependent DNA polymerase (RT). The Retroviridae were originally
classified into three subfamilies: *Oncovirinae* (*onkos*-Greek word meaning 'tumour'), *Lentivirinae* (*lenti*-Latin word meaning 'slow') and *Spumavirinae* (*spuma*-Latin word meaning 'foam'; Matthews, 1979). This family has since been re-classified by disease, tissue tropism, host range, virion morphology and more recently by genome relationships, and is now divided into 7 genera (Table 2). The oncoviruses are the only genera of retroviruses that can transform target cells and are associated with cancers and neurological disease. They are also characterised by the morphology of their core in electron micrographs as type A, B, C and D.

HTLV-I and HTLV-II are members of this genera. Spumaviruses cause a distinct cytopathological effect (vacuolated 'foamy' cytopathology) but do not appear to cause clinical disease. Lentiviruses (slow viruses) characteristically display long periods of latent infection prior to causing immunological and neurological disease. They contain a characteristic cylindrical nucleocapsid core, obvious in electron micrographs, tapering at one end. The genomes of this genera carry a complex combination of regulatory genes in addition to *gag*, *pol* and *env*. The prototype members were visna virus, equine infectious anemia virus and caprine arthritis-encephalitis virus. More recently, human and simian immunodeficiency viruses (HIV and SIV) and the more distantly related feline and bovine immunodeficiency viruses (FIV and BIV) have been classified as members of this genera (Table 2).
## TABLE 2: CLASSIFICATION OF RETROVIRUSES

<table>
<thead>
<tr>
<th>Genus</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avian-leukosis-sarcoma</td>
<td>RSV</td>
</tr>
<tr>
<td>Mammalian C-type</td>
<td>Mo-MLV, FeLV</td>
</tr>
<tr>
<td>B-type viruses</td>
<td>MMTV</td>
</tr>
<tr>
<td>D-type viruses</td>
<td>MPMV</td>
</tr>
<tr>
<td>HTLV-BLV group</td>
<td>HTLV-1 and-2, BLV</td>
</tr>
<tr>
<td>Spumavirus group</td>
<td>HFV, SFV</td>
</tr>
<tr>
<td>Lentivirus group</td>
<td>HIV, SIV</td>
</tr>
</tbody>
</table>

RSV, Rous sarcoma virus; Mo-MLV, Moloney murine leukemia virus; FeLV, Feline leukemia virus; MMTV, Mouse mammary tumor virus; HTLV, Human lymphotropic virus; BLV, Bovine leukemia virus; HFV, Human foamy virus; SFV, Simian foamy virus; HIV, Human immunodeficiency virus; SIV, Simian immunodeficiency virus.
1.2.2 BIOCHEMICAL STRUCTURE OF HIV-1.

The morphology of HIV has been determined by thin section electron microscopy and by negative staining (Christie et al., 1988; Standard et al., 1987). HIV-1 is an enveloped virion 80-100nm in diameter (Matthews, 1979). The lipid bilayer is densely studded with 72 knob like projections (Gelderblom, 1991), derived from a single precursor protein (gp160) encoded by the env gene, of approximately 8nm in diameter (Gelderblom, 1991; Matthews, 1979). Cleavage of gp160 produces gp41 (transmembrane protein-TM) and gp120 (outer surface glycoprotein-SU) which are found in close association on the surface of the virion in the form of a heterodimer. Host cell components have been shown to be present on the viral membrane, such as major histocompatibility complex (MHC) class I or II depending on the infected cell from which the virus buds (Henderson et al., 1987; Arthur et al., 1992; Frank et al., 1996; Fig 2). The morphology within the virion is provided by the protein products of the gag polyprotein gene which form two distinct constituents. Thin section electron micrographs have indicated an electron dense layer underlying and in close proximity to the lipid membrane (Gelderblom et al., 1989). This layer has since been characterised as the matrix (MA) protein encoded by the p17 region of gag and is vital for the structure and integrity of the virion. The second constituent provided by the gag gene is the viral core generated by the p24 region and has a characteristic cone shape prominent in electron micrographs (Gelderblom, 1991). The core contains an RNA genome that
Fig. 2. Biochemical structure of HIV-1 (modified from Fields Virology, 1996).
encodes genetic information required for the replication of HIV. It consists of two identical copies of RNA found in close association with the viral RNA-dependent DNA polymerase (RT), and the nucleocapsid proteins encoded by the gag gene (p6 and p9). The association between the nucleocapsid proteins and RNA genome forms the ribonucleo-protein complex (RNP). The core also contains two other virally encoded enzymes required for the replication of HIV, these are protease (PR) and integrase (IN). There is some evidence to suggest a link between the cone shaped core and the matrix protein. Hoglund et al have proposed a core-envelope link protein attaching the narrow end of the core to the matrix protein (Hoglund, 1990). Nomenclature for the various HIV proteins mentioned has been taken from the revised standardized nomenclature for proteins common to all retroviruses by Leis et al and is summarised in Table 3 (Leis et al., 1988).
TABLE 3: STANDARDIZED NOMENCLATURE FOR HIV PROTEINS

<table>
<thead>
<tr>
<th>HIV protein</th>
<th>Function</th>
<th>Nomenclature</th>
</tr>
</thead>
<tbody>
<tr>
<td>p17</td>
<td>Matrix protein</td>
<td>MA</td>
</tr>
<tr>
<td>p24</td>
<td>Capsid protein</td>
<td>CA</td>
</tr>
<tr>
<td>p9</td>
<td>Nucleocapsid protein</td>
<td>NC</td>
</tr>
<tr>
<td>p10</td>
<td>Protease</td>
<td>PR</td>
</tr>
<tr>
<td>p66/p51</td>
<td>Reverse transcriptase</td>
<td>RT</td>
</tr>
<tr>
<td>p32</td>
<td>Integrase</td>
<td>IN</td>
</tr>
<tr>
<td>gp120</td>
<td>Surface protein</td>
<td>SU</td>
</tr>
<tr>
<td>gp41</td>
<td>Transmembrane protein</td>
<td>TM</td>
</tr>
</tbody>
</table>
1.2.3 GENOMIC ORGANISATION OF HIV-1.

The HIV proviral genome has been well characterised through molecular cloning and extensive sequencing (Fig 3; Hahn et al., 1984). Every HIV particle has two identical positive strands of RNA, each of which is approximately 9.2 kilobases (Kb) long. Common to all retroviruses the HIV genome encodes the three main protein coding genes gag, pol and env flanked by long terminal repeat (LTR) sequences that are required for replication. However, HIV is more complex than conventional retroviruses, such as HTLV-I and HTLV-II, encoding an additional 8 genes; tat, rev, nef, vif, vpr, vpu, vpx and tev. For convenience gag, pol and env have been classified as structural genes, tat, rev and nef as regulatory genes, and vif (virion infectivity factor), vpr (virion protein R), vpu (virion protein U), vpx and tev as accessory genes. In order to facilitate the synthesis of at least 11 distinct proteins, from a genome of less than 10kb, HIV displays great economy in its coding potential employing a complex array of differential RNA splicing and overlapping translational reading frames.

1.2.4 STRUCTURAL GENES OF HIV-1.

Transcription of the HIV genome is characterised by a transition from the synthesis of short multiply spliced messenger RNAs (mRNAs; approximately 2kb), which encode the regulatory proteins early in infection, to the production of larger singly spliced (4.5kb) and full length unspliced (9.2Kb) mRNAs encoding the
Fig. 3. Genomic organization of HIV-1 (modified from Fields Virology, 1996).
structural proteins of HIV. The transition from early gene expression to late gene expression is controlled by the viral trans-activator proteins tat and rev. The function of these two regulatory proteins will be discussed more fully in section 1.2.5.

The gag gene codes for the main non-glycosylated structural components of the virus particle. Initially a polyprotein precursor (p55) is translated from full length mRNA and is subsequently cleaved by the viral protease to produce four distinct gag proteins; p17 (matrix protein), p24 (major capsid protein), p9 (nucleic acid binding protein) and p6 (proline-rich protein). Various domains of this gene play pivotal roles in the assembly and release of virus particles (Mervis et al., 1988) and are discussed more fully in section 1.2.8.6.

The pol gene codes for several enzyme activities required during the life cycle of HIV. To date the pol gene of all retroviruses have been shown to be expressed initially as a gag:pol fusion protein (Weiss et al., 1985). In HIV the gag and pol genes overlap by 241 nucleotides and the pol gene is in a -1 reading frame with respect to gag (Vaishnav et al., 1991). Expression of the pol gene is facilitated by a ribosomal frameshift which occurs at a low frequency (5%) and is directed by a short homopolymeric sequence located in the overlap between the gag and pol open reading frames that allows a 'slip back' of the ribosome into the -1 position (Jacks et al., 1988). This results in the production of large quantities of the gag structural genes and relatively small quantities of the pol gene products. This p160 gag:pol polyprotein precursor is cleaved by the viral protease to produce three distinct enzymes; p10 (PR), p66/p51 (RT) and p32 (IN) and four distinct gag
proteins (mentioned above).

The env gene codes for the viral glycoproteins which are produced by cleavage of the gp160 polyprotein precursor by a cellular protease in the Golgi. Two glycoproteins are produced, gp120 and gp41, which are located on the surface of the virion embedded into the plasma membrane. gp120 is an external glycoprotein (gp), whereas gp41 is a transmembrane glycoprotein anchoring the envelope complex to the virus. Both gp120 and gp41 are extensively glycosylated and contain complex N-linked carbohydrates. gp120 is thought to be associated with tissue tropism of the virus while gp41 is thought to mediate fusion of the virus and cellular membranes prior to entry. gp120 and gp41 are non-covalently associated and are thought to form tetramers composed of dimers of each glycoprotein (Vaishnav et al., 1991). The envelope protein, gp120, consists of five hypervariable regions (V1-V5) interspersed with five more conserved regions (C1-C5; Modrow et al., 1987; Fig 4). Leonard et al., have determined the location of 9 disulphide bonds in a prototypic gp120 and have proposed a secondary structural model in which gp120 is composed of five major cysteine loops (Leonard et al., 1990).

1.2.5 REGULATORY GENES OF HIV-1.

The tat gene is encoded by two exons, one of which is found 5' of the env gene and the other is found within the coding region of the env gene (Fig 3; Karn et al., 1991). The tat gene encodes a 14Kd protein that is essential for viral
Fig. 4. Secondary structural model of gp120 (modified from Leonard et al., 1990).

\( \gamma \), high mannose or hybrid-type oligosaccharide structures. \( \psi \), complex-type oligosaccharide structures. Hypervariable regions 1 to 5 are indicated by enclosed boxes.
replication. This nuclear protein is a transactivator of LTR-directed gene expression and binds to a nascent RNA structure located within the LTR known as the trans-activation response element (TAR; Sodroski et al., 1985; Vaishnav et al., 1991). Tat can act at multiple levels in the viral life cycle although its main function is to increase the level of transcripts from genes linked to the HIV-1 LTR. However, it has also been shown to function at the post-transcriptional level and has recently been shown to decrease the activity of the MHC class 1 gene promotor (Howcraft et al., 1993; Felber et al., 1993).

The rev gene is also encoded by two exons, one is found 5' of the env gene and the other, as with the tat gene, is found within the coding region of the env gene (Fig 3; Karn et al., 1991). The rev gene encodes a 19Kd nuclear protein, which binds to a specific RNA target sequence, found within the env coding region, known as the rev responsive element (RRE; Arrigo et al., 1989). The rev protein appears to affect the nuclear-cytoplasmic distribution of HIV RNAs post-transcriptionally by either inhibiting spliceosome function or by activating the export of the singly spliced (4.5Kd) and unspliced (9.2Kd) HIV RNAs from the nucleus to the cytoplasm. This function promotes the expression of the larger structural proteins required for virus assembly and maturation.

The nef gene is encoded by a single exon which is located at the 3' end of the genome overlapping the 3' end of the env gene and the 3' LTR (Fig 3). The nef gene encodes a 25-27Kd protein (Allan et al., 1985), which is found in association with the inner membrane courtesy of post-translational myristoylation (Kan et al., 1986). The function of this regulatory protein is somewhat uncertain. It
was originally thought to act as a negative regulatory factor inhibiting HIV replication (Niederman et al., 1989; Ahmad et al., 1988). However, more recently it has been suggested that this protein may play a role in disease pathogenesis (Schwartz et al., 1995; Miller et al., 1994; Schwartz et al., 1995; Kestler et al., 1991). It has also been found that expression of this protein is inversely linked to the expression of the cell surface CD4 receptor (Garcia et al., 1991) indicating it may act to downregulate CD4 expression of infected cells.

1.2.6 ACCESSORY GENES OF HIV-1.

The vif gene is encoded by a single exon overlapping the pol gene and encodes a 23Kd protein (Fig 3; Kan et al., 1986). The function of this protein is uncertain although there is some evidence to indicate that it may promote infectivity of cell free virus since vif negative mutants show a profound reduction in their infectivity (Fisher et al., 1987). Although the exact mechanism by which vif affects viral infectivity has not been determined it is thought this protein acts during the late phases of viral replication (Kan et al., 1986; Fisher et al., 1987).

The vpu gene is located at the 5’ end of the env gene and encodes a 15-20Kd protein (Fig 3; Cohen et al., 1989). vpu is unique to the HIV-1/SIV<sub>cpz</sub> group of primate lentiviruses and is not found in either HIV-2 or other SIV isolates. Mutational analyses using vpu negative isolates have suggested that this protein may be involved in facilitating assembly and/or release of virus particles (Klimkait et al., 1991; Strebel et al., 1989).
The vpr gene is encoded by a single exon which overlaps the vif gene and encodes a 15Kd protein (Fig 3). This protein accelerates the replication and cytopathic effects of HIV and appears to act at an early step in the replication cycle (Ogawa et al., 1989). Unlike onco-retroviruses HIV does not require mitosis for entry of viral nucleic acid into the host cell nucleus (Bukrinsky et al., 1992). It has been suggested that vpr, in association with MA protein from the gag region, may be crucial in the translocation of the pre-integration complex into the nucleus of a non-dividing cell (Freed et al., 1995; Heinzinger et al., 1994). A role for vpr in the regulation of cellular functions has also been suggested. It has been demonstrated that vpr causes terminal differentiation of rhabdomyosarcoma cells (Levy et al., 1993), and expression of vpr in the yeast Saccharomyces cereviseae resulted in cell growth arrest and gross cell enlargement (Macreadie et al., 1995). Hence, vpr may function to bring about growth arrest to enable genomic integration to occur.

The vpx gene has, to date, only been found in HIV-2 and SIV isolates but not in HIV-1. It encodes a 14Kd protein (SIV) and a 16Kd protein (HIV-2; Yu et al., 1988; Franchini et al., 1988). It has been suggested that the vpx gene arose from a duplication in the vpr gene (Tristem et al., 1992). A number of studies have suggested that vpx is required for efficient replication in PBMCs (Guyader et al., 1989; Kappes et al., 1991; Yu et al., 1991). One such study showed that mutant vpx negative isolates infect primary lymphocytes with a reduced efficiency (Guyader et al., 1989). Also mutations in two adjacent regions of this gene have been shown to disrupt the incorporation of this protein into the mature virion (Park
et al., 1995).

The tev gene encodes a 28Kd protein and is a hybrid protein expressed by a novel multiply spliced mRNA. This mRNA begins with the first exon of tat continues through env and ends in the second exon of rev. The function of this protein is not clear, although it has been shown to contain both tat and rev properties. Expression is decreased in the presence of rev suggesting it is produced early on in the virus life cycle and may play a regulatory role in the initial stages of virus expression (Benko et al., 1990).

1.2.7 ORIGIN AND EVOLUTION OF HIV-1.

Primate lentiviruses include HIV-1, HIV-2 and a number of viruses isolated from nonhuman primates collectively known as simian immunodeficiency viruses (SIV). Extensive genetic analysis has been carried out on these primate lentiviruses and they have been found to be more closely related to each other than they are to lentiviruses of other mammals (Myers et al., 1992b). In order to further classify these primate lentiviruses phylogenetic trees have been constructed, derived from pol protein sequences, which have identified five distinct groups (Sharp et al., 1994). One group contains viruses present in humans (HIV-1) and chimpanzees, a second group contains viruses found in humans (HIV-2), macaques and sooty mangabeys, and the final three groups consist of viruses currently found within single species (African green monkey, the mandrill and Sykes’ monkey).

Theories of the origin of HIV viruses have remained extremely
controversial. Serological studies on nonhuman primates have been carried out to identify animals with antibodies cross-reactive with HIV or closely related viruses. Seropositive nonhuman primates include African green monkeys, mandrills and sooty mangabeys (Kanki et al., 1985). Similar serological studies carried out in humans from West Africa have suggested infection with a virus closely related to SIV (Barin et al., 1985). This virus, designated HIV-2, was isolated by Guyader et al., who found it to be closely related to a virus isolated from a captive macaque (SIVmac), but significantly different from HIV-1 (Guyader et al., 1987). This finding was somewhat puzzling since the natural habitat for macaques is Asia not Africa, and Asian macaques have not been found to be infected with SIV in the wild.

Hirsch et al., suggested that macaques could probably have been infected with SIV through a cross species transmission from an African primate (Hirsch et al., 1989). Epidemiological data point to sooty mangabeys as the source of SIVmac. There would therefore appear to be good phylogenetic evidence to suggest that the sooty mangabey is the natural reservoir for HIV-2 and that infection of man probably represents a cross species transfer.

The natural reservoir for HIV-1 remains unclear, although it is highly likely that known strains of HIV-1 have arisen in a similar fashion to HIV-2, namely following a cross species transmission(s). Phylogenetic analysis has shown a close relationship between HIV-1 and SIV infecting chimpanzees (SIVcpz), which may indicate that the natural reservoir for HIV-1 is in fact the chimpanzee (Peeters et al., 1992; Huet et al., 1990). However, in a seroepidemiological study of 94 chimpanzees (originating from Gabon, Zaire and Cote d'Ivoire) only three animals
were found to be infected (Peeters et al., 1992), casting doubt on these primates being the natural host. It is therefore possible that both humans and chimpanzees have been infected following cross species transmission from a third, yet unidentified primate. Gojobori et al., have constructed phylogenetic trees for a number of primate lentiviruses, comparing nucleotide sequences from five regions of the genome (LTR, gag, pol, env, nef; Gojobori et al., 1990a). They analysed SIV_{agm} and SIV_{mnd} and found that these isolates grouped with either HIV-1 or HIV-2 depending on the gene analysed. For example, analysis of the env region suggests SIV_{agm} is closely related to HIV-2, while analysis of the nef region suggested SIV_{agm} is more closely related to the HIV-1 group. These results suggest a recombination event may have occurred between simian viruses prior to the emergence of HIV-1 and HIV-2. It is therefore possible that HIV-1 and HIV-2 have evolved from two separate cross species transfers from non human primates to humans, with one occurring in Central Africa, possibly from chimpanzees giving rise to HIV-1 and a second in Western Africa most probably from sooty mangabeys giving rise to HIV-2.

Divergence times between HIV-1, HIV-2 and SIVs have been estimated by a number of groups. A comparison between the complete genome of SIV_{agm} (isolated from an African green monkey) and known HIV and SIV sequences suggested that HIVs and SIVs have "diverged gradually in concert with the evolution of primates" and that HIV-1 and HIV-2 have been present in the human population for some time, having diverged during human evolution (Fukasawa et al., 1988). This implies a diversification time of millions of years. For more closely
related viruses Hircsh et al., suggested that SIV_{sm}/SIV_{mac} may have diverged from HIV-2 approximately 30 years ago (assuming a mutation rate for the gag gene of 0.5% per virus per year; Hirsch et al., 1989), similar to other estimates (Sharp et al., 1988). Extension of this investigation for the divergence of HIV-1 from HIV-2 predicts a time of hundreds of years. Similarly, Yokoyama et al., estimated HIV-1 and HIV-2 diverged about 280 years ago, assuming the pol gene of lentiviruses evolved at the same rate as viral oncogenes (0.5 x 10^{-3} nonsynonymous substitutions per site per year; Yokoyama, 1988). Yet another study estimates that HIV-1, HIV-2 and SIV_{agm} all diverged approximately 150 years ago (Sharp et al., 1988), although in this study the rate of evolution of the pol gene was estimated to be 0.96 x 10^{-3}. All these studies are inconsistent with the proposal that HIV-1 and HIV-2 diverged from a common ancestor as recently as 40 years ago (Smith et al., 1988). The range in estimated times of divergence illustrates the weakness of this aspect of evolutionary analysis in reconstructing convincing histories of virus evolution.

Attempts have been made to further classify the five groups of primate lentiviruses. This is especially important within the two lineages containing human viruses since development of a successful vaccine will require knowledge of the extent of HIV genetic diversity worldwide. Phylogenetic analysis of HIV-1 strains has produced a classification system in which different subtypes or clades have been distinguished. The first attempt at describing HIV-1 subtypes was accomplished by Myers et al., using env gene sequences (Myers et al., 1992a). This analysis provided evidence of at least 6 distinct env subtypes, designated A to
F. Subtype A is representative of sequences from Central Africa (Zaire, Uganda and Rwanda; Myers et al., 1991). Subtype B is found most prevalent in North and South America and Western Europe although subtype B sequences have been detected in most continents (Pfeifer et al., 1991). Subtype C includes isolates from South Africa (Myers et al., 1992a), India (Dietrich et al., 1993), Zambia (McCutchan et al., 1992b) and Djibouti (Louwagie et al., 1993). Subtype D includes isolates from Central Africa (Myers et al., 1991). Subtype E isolates are mainly found in Thailand although isolates from Central Africa (McCutchan et al., 1992a) and India (Grez et al., 1994) have been described. Subtype F includes viruses from Brazil, Romania and Zaire (Louwagie et al., 1993; Dumitrescu et al., 1994; Potts et al., 1993; Louwagie et al., 1994). Among subtypes A to F, viruses that belong to different subtypes exhibit up to 30% variation in their env coding sequences.

An independent study by Louwagie et al., examined the diversity of HIV-1 among 70 sequenced isolates from 15 different countries using a different region of the genome, the gag gene (Louwagie et al., 1993). This analysis described 7 distinct subtypes originally denoted A to G. Unfortunately when Louwagie et al., characterised the three new subtypes, E to G, only four env subtypes had been proposed. However by the time this gag data had been published a fifth new env subtype had been identified and named 'E' on the basis of the established nomenclature (Myers et al., 1992a). However, gag E and env E isolates were found to belong to distinct viral groups (Louwagie et al., 1994), and so gag E to G had to be renamed to gag F to H, illustrating the problems which may arise during the
naming of new subtype sequences. The subtype of most HIV-1 isolates have been found to be congruent for gag and env sequences, although some discrepant sequences are apparent. For example, the African isolates (MAL, from Zaire), have an env sequence similar to subtype D viruses but a gag sequence similar to subtype A (Alizon et al., 1986). Also, viruses characterized from Thailand (CM238 and CM243), have a subtype E env sequence but a subtype A gag sequence (Louwagie et al., 1993). This switching of subtype between different regions of the genome suggests that some subtypes may have arisen through recombination between two divergent viruses. More recently, four additional sequence subtypes have been assigned on the basis of their env sequences, termed G, H, I and J. These include viruses from Gabon, Cameroon and Zaire (Jansens et al., 1994). Collectively subtypes A to J are classified as group M (major subtypes).

Recently much more diverse variants of HIV-1 have been identified. Two independent studies have isolated HIV-1 isolates which, although they are more closely related to HIV-1 than any of the other four groups of primate lentiviruses, are considerably more diverse than the previously described subtypes of HIV-1. De Leys et al., report a novel isolate obtained from two individuals, originally from Cameroon, which exhibit major differences in immunological and biological properties and in the U3 region of the viral LTR. This isolate has been designated ANT 70 (Deleys et al., 1990). Gurtler et al., also report a novel isolate from a Cameroonian AIDS patient with a V3 configuration which has not yet been described for other HIV strains. This isolate has been designated MVP-1580 (Gurtler et al., 1994). Both novel isolates, ANT-70 and MVP-1580, are considered
to constitute a new subtype, group O (meaning outlier; Peeters et al., 1996). Although these two isolates have been grouped together they are as divergent from each other as subtypes A and B and could therefore represent separate subtypes themselves. No doubt as more novel isolates are recovered this may be resolved.

A highly divergent HIV-1 related virus (SIV<sub>cpz-ant</sub>) was isolated from a wild captured chimpanzee originating from Zaire (Vanden-Haesevelde et al., 1996). Phylogenetic analysis using regions from gag and pol revealed that this isolate was more closely related to HIV-1 isolates and other chimpanzee isolates than to the other four major lineages. However, this isolate grouped outwith this lineage being no more similar to SIV<sub>cpz-gab</sub> than to HIV-1 isolates. The isolation of this SIV isolate is important regarding the origin(s) of HIV-1 and SIV<sub>cpz</sub> viruses. The phylogenetic position of SIV<sub>cpz-ant</sub> suggests that the two major HIV-1 clusters, groups M and O, must have arisen through two independent transmission events. Whether this cross-species transmission occurred from chimpanzee to human is unknown, although the low seroprevalence among chimpanzees argues against it. Therefore, it is possible that both humans and chimpanzees have acquired their lentiviruses from a third, as yet unidentified, African primate species.

1.2.8 LIFE CYCLE OF HIV-1.

The life cycle of HIV-1 requires the infection of a human cell (Fig 5). It can be divided into a number of phases summarised below. (1) Prior to infection an interaction with a receptor molecule occurs facilitating the attachment of the virus to the cell surface. The high affinity receptor for the virus has been shown to be
Fig. 5. Life cycle of HIV-1 (modified from Fields Virology, 1996).
the CD4 molecule found on the surface of a number of human cells. (2) Following attachment the virus is internalized and uncoated. The mechanism of entry is somewhat controversial and a number of mechanisms have been proposed, including pH dependent and independent processes. (3) Once internalized, the genomic RNA is transcribed into DNA using the viral enzyme RT. (4) The proviral DNA is then integrated into the host chromosomal DNA utilizing another virally encode enzyme IN. (5) A latent phase may follow viral integration restricting the life cycle until the infected cell is activated allowing the transcription of viral genomic RNA and mRNA. (6) Protein synthesis and processing then occur followed by virus assembly and maturation as the virion buds from the cell surface acquiring its coat.

1.2.8.1 ATTACHMENT.

The first step in the initiation of infection is the binding of the virus to the surface of a susceptible cell via an interaction with a specific receptor. In the case of HIV this specific receptor was found to be dependent on the surface presentation of cluster determinant 4 (CD4; Maddon et al., 1986; McDougal et al., 1986; Dalgleish et al., 1984; Klatzmann et al., 1984). The CD4 molecule is a transmembrane glycoprotein of 58Kd and consists of a extracellular region, transmembrane region and a cytoplasmic region at the C-terminal. The extracellular region is folded into four domains, D1 to D4 (Maddon et al., 1985). CD4 is a member of the immunoglobulin superfamily and is present predominantly on T
helper cells but can also be found on B cells, cells of the monocyte-macrophage lineage and specialized cells of the CNS (Maddon et al., 1986). The CD4 molecule interacts with the envelope glycoprotein gp120 prior to infection. A number of studies have been carried out which have precisely mapped the regions on both the CD4 molecule and gp120 involved in this highly specific interaction (Clayton et al., 1989; Arthos et al., 1989; Lasky et al., 1987; Kowalski et al., 1987; McDougal et al., 1986). The carboxy-terminal of gp120 is thought to interact with a region located in the first domain of the CD4 molecule homologous to the second complementary determining region (CDR2) of immunoglobulin variable region. This interaction is also thought to facilitate the fusion of the virus membrane with the cell membrane and will be discussed in section 1.2.8.2.

Chemokine receptors have been found to act as secondary receptors for HIV-1 infection. A member of the α chemokine receptor family was the first to be identified which acts as a co-receptor for T cell line tropic strains (Feng et al., 1996). This receptor was previously named LESTR/fusin and has recently been redesignated CXCR-4 (Bleul et al., 1996). Subsequent studies revealed that a member of the β chemokine receptor family, CCR-5, serves as a co-receptor for macrophage tropic isolates or dual tropic primary isolates (Alkhatib et al., 1996; Deng et al., 1996; Dragic et al., 1996). Bruel et al., have shown that infection of HeLa-CD4+ cells, PBMCs and CXCR-4 transfectants were inhibited by the CXC chemokine stromal cell derived factor-1 (SDF-1), the natural ligand for CXCR-4, but did not affect CCR-5 mediated infection of macrophage tropic or dual tropic primary HIV-1 isolates (Bleul et al., 1996). Similar inhibition studies have revealed
that RANTES, MIP-1α and MIP-1β (natural ligands for CCR-5) blocked infection of lymphocytes but failed to block infection of macrophages (Cheng-Mayer et al., 1997; Schmidtmayerova et al., 1996). These findings suggest that HIV-1 variants may utilise different co-receptors for infection of macrophages. A further difficulty with understanding the tropism of HIV for macrophages has arisen from the observation that macrophages, common with most human cells in vivo, express the CXCR-4 receptor but are non-permissive for infection by SI strains that specifically use this second receptor. A recent study by Rana et al., found that primary CD4+ T cells recovered from individuals homozygous for a mutation in the CCR-5 chemokine receptor (see below) were resistant to infection with macrophage tropic strains but permissive to infection with T cell tropic and dual tropic isolates that use CXCR-4, CCR-5, CCR-3 or CCR-2b (Rana et al., 1997). Clearly expression of CD4 and CXCR-4 in macrophages is insufficient to confer susceptibility. However, transfection of these genes into non-human cells such as cat, hamster and mink was sufficient to permit infection with SI isolates.

There is little information on the distribution of chemokine receptors amongst different cell types in vivo. However, CXCR-4 is thought to be widely distributed in human cells and tissues (Feng et al., 1996) and has been detected in PBMCs and a number of CD4+ cell lines (Loetscher et al., 1994). CCR-5 has been found to be expressed in primary monocytes/macrophages, primary T cells and granulocyte precursors (Alkhatib et al., 1996; Deng et al., 1996), while CCR-3 has so far only been detected in eosinophils (Daugherty, 1996) and microglia (He et al., 1997).
Differential chemokine secondary receptor usage of T cell tropic and macrophage tropic variants suggests that the cellular host range of HIV-1 may be determined by utilization of co-receptors. Recent studies have shown that the V3 region is a determinant of chemokine receptor usage (Cocchi et al., 1996; Choe et al., 1996). Choe et al., demonstrated that recombinant viruses containing chimaeric glycoproteins with a V3 loop from primary macrophage tropic isolates were able to infect HeLa-CD4 cells more efficiently when either CCR-3 or CCR-5 was exposed on the target cell (Choe et al., 1996). Wu et al., and Trkola et al., have shown that binding of MIP-1β to CCR-5 was blocked by the gp120/CD4 complex, suggesting gp120 specifically interacts with this co-receptor (Wu et al., 1996; Trkola et al., 1996). The gp120-CCR-5 interaction was enhanced by CD4 suggesting an interaction between gp120 and CD4 may result in a conformational change within gp120 facilitating the binding of CCR-5. These findings are consistent with previous observations that the V3 region is an important determinant of HIV-1 tropism and phenotype (see section 1.4.3).

Secondary receptor usage has also been shown to influence infection of microglia in brain. Recently, He et al., reported that macrophage tropic HIV-1 viruses that use CCR-5 and CCR-3 as co-receptors were able to efficiently infect microglia but that a T cell tropic virus that uses CXCR-4 was not (He et al., 1997). These findings suggest that both CCR-3 and CCR-5 serve as co-receptors for infection of brain microglia. This is supported by the observation that CCR-5 and CCR-3 ligands can inhibit infection of microglia by several primary neurotropic HIV-1 isolates (He et al., 1997). This is an interesting finding and may help to
explain the overlap observed between macrophage tropism and that of microglia since blood derived monocyte/macrophages express CCR-5 but do not express CCR-3.

The discovery of co-receptors for HIV-1 infection also has important implications regarding disease pathogenesis. Two recent studies have identified a mutant allele of CCR-5 that is homozygous in two exposed, yet uninfected individuals (Liu et al., 1996; Samson et al., 1996). This allele was found to contain a 32 base-pair deletion, resulting in a truncated protein which was unable to mediate chemokine signalling. These findings have far reaching implications suggesting CCR-5 is a major determinant of HIV-1 infection in vivo and that the absence of CCR-5 may confer protection against HIV-1 infection. Consistent with this hypothesis, Samson et al., reported that out of 723 HIV-1 infected individuals none were found to harbour this mutant allele while 8 out of 704 uninfected individuals were found to express this allele (Samson et al., 1996). They also found a reduced frequency of heterozygotes in the HIV-1 infected group compared with the uninfected group suggesting a single copy of this mutant allele may confer some degree of protection. Hwang et al., reported evidence for an association between the heterozygous genotype with a slower rate of CD4+ T cell decline and a lower viral load (Huang et al., 1996). These studies were fairly small and more extensive analysis of the populations will be required to determine the true relevance of these observations in disease pathogenesis.

The observation of some degree of protection awarded by the presence of CCR-5 mutants is potentially important in viral transmission. Mutant CCR-5 alleles
may confer resistance to infection following sexual transmission since macrophage cells are likely to be the first cell type encountered by the virus upon transmission. This would be consistent with previous reports that macrophage tropic isolates are predominantly detected following transmission (Zhu et al., 1993). However, further epidemiological studies will be required to examine the distribution of mutant alleles in other risk groups to determine if this is an important factor during transmission and if this mutation alone confers natural resistance to infection with HIV-1. Indeed, a combination of genotypic variation within virions and co-receptor usage may determine the transmission of HIV-1 variants.

1.2.8.2 FUSION AND ENTRY.

Enveloped animal viruses, such as HIV, enter host cells following fusion with the cell membrane. The mode of entry of enveloped viruses can be broadly divided into two mechanisms. The first is by a pH-dependent process involving internalization of the virus by receptor mediated endocytosis (RME) into acidic compartments (endosomes), where a reduction in pH induces a conformational change in the TM glycoprotein, exposing the hydrophobic fusion domain thereby facilitating fusion of viral and endosomal membranes. The second possible mode of entry is via a pH independent process. This mechanism involves the direct fusion of the virus envelope with the plasma membrane of the cell and is independent of a decrease in pH. The mechanism by which HIV enters cells is one of the most poorly understood aspects of the virus life cycle. Thin section EM studies on viral
fusion and uptake have provided evidence for both RME (via clathrin coated pits) and direct fusion with the cell membrane (Pauza et al., 1988; Stein et al., 1987; Grewe et al., 1990). Early studies on the fusion and uptake of HIV concentrated on determining whether uptake was pH dependent or independent. To enable this distinction to be made a number of studies have been carried out using various methods. HIV infection was reported to be inhibited by treating cells with lysosomotropic agents (weak bases) such as ammonium chloride and amantadine (Maddon et al., 1986). Subsequently, however, it was reported that lysosomotropic agents and carboxylic ionophores, such as monensin, did not abrogate HIV fusion and infectivity (McClure et al., 1988; Stein et al., 1987). Following these studies Maddon et al constructed cell lines expressing defective CD4 molecules that were unable to undergo RME, yet found they remained susceptible to HIV infection. This suggests a mechanism whereby binding of the virus to the CD4 molecule is followed by the direct introduction of its RNA into the cell by fusion with the cell membrane (Maddon et al., 1988).

Little more is known about the fusion event itself, how the fusion process is triggered and the intra- and intermolecular associations involved. It is likely that this step is mediated by a region of hydrophobic amino acids at the amino terminal end of the TM protein, gp41 (Kowalski et al., 1987; Freed et al., 1990). Substitutions within gp41 which insert charged amino acids into this hydrophobic region have been found to impair syncytium formation (Kowalski et al., 1991). Regions in the carboxy and amino terminal 30 residues of gp120 have been found in association with gp41 (Ivey Hoyle et al., 1991; Helseth et al., 1991), and are
thought to form a molecular pocket masking the amphipathic regions of gp41. Interaction of gp120 with the cell surface receptor CD4 is subsequently thought to trigger conformational changes resulting in dissociation from gp41 exposing the fusogenic domain (Freed et al., 1990). This also results in the shedding of gp120 since both glycoproteins are weakly associated through non-covalent bonds on the virion surface (Gelderblom et al., 1987). Epitope mapping studies have shown that upon binding of soluble CD4 to gp120 conformational changes important for fusion occur in the V3 loop (Sattentau et al., 1991). Similarly, neutralizing antibodies against V3 can inhibit HIV infection without preventing CD4 binding, suggesting the V3 loop is involved in post-binding events (Skinner et al., 1988).

1.2.8.3 REVERSE TRANSCRIPTION.

All retroviruses must reverse what is generally thought to be the normal flow of genetic information, DNA to RNA to protein, because their genome is entirely composed of RNA and hence must convert genomic RNA into DNA prior to protein synthesis. This phenomenon is achieved in the cell cytoplasm by the action of the virally encoded enzyme RT. The viral RNA genome is synthesised using a host DNA-dependant RNA polymerase (DNA-dep RNA pol) and as a consequence contains cellular post-translational modifications, namely the 5' end of the genome is capped (Gppp) and the 3' end has a poly A tail (AAA; Weiss et al., 1985). Reverse transcription is initiated at the 5' end of the RNA genome and requires a host cell molecule, transfer RNA (tRNA), to act as the initial primer of
DNA synthesis. This tRNA is complementary to a region of the genome known as the primer binding site (PBS) near the 5' end of the viral genome. An RNA-dependant DNA polymerase (RNA-dep DNA pol) activity of RT uses this primer to synthesise a DNA copy of the 5' U5 and R regions of the genome. A second component of RT, RNaseH, then degrades the RNA copied by the polymerase allowing the newly synthesised DNA region to act as a second primer binding site to the complementary R sequence at the 3' end of the genome. This transfer is known as the first jump. Synthesis of the (-) DNA strand is then completed using the RNA-dep DNA pol, while the RNaseH degrades the remainder of the RNA template. A small region is left upstream of the 3' U3 and R regions which acts as the primer for (+) strand DNA synthesis. Synthesis of the (+) strand begins at the 3' end of the (-) DNA strand and continues through U5 and the tRNA primer. A second jump occurs, as the tRNA primer is complementary to the 5' PBS, completing the synthesis of the (+) DNA strand. Synthesis of the (+) DNA strand can occur as an intermolecular or an intramolecular event.

1.2.8.4 INTEGRATION.

Following reverse transcription double stranded HIV DNA is transported to the nucleus in association with a nucleoprotein complex. Once in the nucleus HIV is integrated into the host genome in its linear form. Integration is catalyzed using the virally encoded enzyme IN. This process can be separated into three stages, the first being the removal of two bases from the 3' ends of both DNA strands.
Second, the 3' ends of the viral DNA are joined in a concerted reaction to previously nicked sites in the host DNA. Finally, the gaps in the mismatched intermediate are repaired (Engelman et al., 1991). The integrated HIV DNA genome is known as the HIV provirus. This integration process was thought to occur at random sites within the host genome however there is some evidence that integration occurs at 'hot spots' in the host genome (Scherdin et al., 1990).

Following integration some retroviruses persist for the lifetime of the cell, and may spread further amongst the descendants of the cell as each harbour copies of the original provirus in their genome.

1.2.8.5 SYNTHESIS OF HIV-1 RNA.

Synthesis of HIV RNA is initiated by cellular proteins called transcription factors. The 5' LTR of HIV contains similar sequences to regions of the host genome which bind these proteins initiating transcription. The HIV provirus can therefore effectively hijack the cellular transcription machinery. As previously mentioned transcription occurs in two stages, early and late. Cellular transcriptional factors initiate mRNA synthesis and maintain it at a basal level which is insufficient to drive replication. Significant levels of fully spliced HIV mRNA are achieved by the action of the regulatory protein tat. This protein binds a specific region at the 5' end of the mRNA termed trans-activation response element (TAR). It is thought that tat acts to stimulate elongation by RNA polymerase II leading to the synthesis of complete mRNAs. When sufficient mRNA is produced rev acts as
a balance shifting the synthesis of short, multiply spliced mRNAs to the production of late, singly spliced and unspliced mRNAs encoding the structural proteins.

1.2.8.6 ASSEMBLY AND RELEASE.

Retroviral assembly and release is controlled by interactions between the gag proteins, viral genomic RNA and the plasma membrane of the cell. The pol and gag gene products are incorporated into virions in the form of their polyprotein precursors, p160 (gag:pol) and p55 (gag). They undergo proteolytic cleavage during or after budding from the cellular membrane to form mature virus particles. The amino-terminus of these precursors are post-translationally modified by the addition of myristic acid which facilitates the interaction with the membrane (Vazeux et al., 1992). The gag and gag:pol precursors are cleaved by the viral PR enabling morphological maturation of virions. Incorporation of envelope glycoproteins during virus budding is essential for the formation of infectious virions. HIV envelope glycoproteins are incorporated as a gp120-gp41 complex which mediates attachment and membrane fusion. The process by which this complex is incorporated into virions is not well understood. However, a number of studies suggest that the MA protein and the cytoplasmic tail of gp41 may interact during virus assembly and release (Freed et al., 1996; Dorfman et al., 1994). Finally the virion buds from the host cell membrane simultaneously acquiring a lipid bilayer envelope.
1.2.9 THERAPEUTIC STRATEGIES.

1.2.9.1 ANTI-RETROVIRAL AGENTS.

The quest to abrogate HIV infection has proved to be a very difficult task. The complex nature of this virus and its life cycle has impeded and mystified scientists in this field for almost 15 years. A number of strategies have been employed, not only to try and prevent infection and replication of this elusive virus, but also to prolong and improve the lives of those infected. A great deal of effort has been directed towards the identification of viral receptor(s) and elucidating the life cycle of HIV since its discovery in 1982. An important impetus for such investigations is the possibility of the development of effective antiviral agents. A number of potential targets, during the life cycle of HIV-1, have been identified which may present opportunities for antiviral therapy (Table 4). Reverse transcription is a phenomenon unique to retroviruses. The first successful inhibitor of RT activity was a dideoxynucleoside analogue that had initially been synthesized as a potential therapy for cancer, 3'-azido-3-deoxythymidine or AZT (zidovudine). It acts as a competitor for nucleotides used by the polymerase and is a potent chain terminator of viral DNA synthesis. Initial studies on the efficacy of this antiviral drug demonstrated increased levels of CD4+ T cell counts, and delayed development of AIDS-defining symptoms in symptomatic and asymptomatic individuals (Volberding et al., 1995; Fischl et al., 1990; Fischl et al., 1987). There has been a great deal of debate regarding when, during infection, AZT should be
TABLE 4: POTENTIAL TARGETS FOR THERAPEUTIC DRUGS.

<table>
<thead>
<tr>
<th>Steps in viral replication</th>
<th>Identified drugs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attachment</td>
<td>sCD4, CD4-peptides, CD4-Ig</td>
</tr>
<tr>
<td>Reverse transcription</td>
<td>AZT, ddI, ddC, stavudine</td>
</tr>
<tr>
<td>Integration</td>
<td>none</td>
</tr>
<tr>
<td>Viral replication</td>
<td>saquinavir, idinavir, ritonavir, nelfinavir</td>
</tr>
<tr>
<td>Protein synthesis</td>
<td>Ribozymes</td>
</tr>
<tr>
<td>Virion assembly</td>
<td>Myristic acid analogues</td>
</tr>
<tr>
<td>Virion release</td>
<td>INF-α</td>
</tr>
</tbody>
</table>

Ig, immunoglobulin; AZT, 3’-azido-3-deoxythymidine; ddI, 2’,3’-dideoxyinosine; ddC, 2’,3’-dideoxycytidine.
administered to provide the greatest benefit. A number of early studies demonstrated that early administration was more effective in slowing progression to disease (Fauci, 1993a; Cooper et al., 1993a; Graham et al., 1992a; Volberding et al., 1995a). However, the Concorde trial revealed that patients who received AZT prior to developing symptoms progressed to AIDS as quickly as those who were given AZT once symptoms had developed (Concorde coordinating committee. 1994). The benefits of AZT monotherapy have been shown to be of limited duration with resistant strains emerging, especially after prolonged therapy, associated with mutations at four positions in HIV RT (Boucher et al., 1992).

Other nucleoside analogues that inhibit RT activity have also been developed, didanosine (2',3'-dideoxyinosine [ddI]; Lambert et al., 1990), zalcitabine (2',3'-dideoxycytidine [ddC]; Yarchoan et al., 1994; Yarchoan et al., 1988) and stavudine (d4T; Broune et al., 1993; Anonymous., 1993). Combinational therapy with one or more nucleoside analogue drug may reduce the emergence of resistance through synergistic or additive effects (Fischl et al., 1995; Knox et al., 1996; Larder et al., 1996).

The limited success of nucleoside analogue RT inhibitors (AZT, zalcitabine, ddI, ddC) in HIV infected patients led to the development of drugs that use alternative viral targets. The protease of HIV is essential for replication of the virus. It cleaves the gag (p55) and gag:pol (p160) polyproteins into the structural gag proteins and three virally encoded enzymes, RT, IN and PR itself. HIV protease is also distinct from human proteases making it a very favourable target for antiviral therapy. To date four protease inhibitors have shown antiviral activity
in patients; saquinavir, idinavir, ritonavir and nelfinavir (Nelson, 1996; Vella, 1994). It was initially thought that due to the nature of this enzyme the development of resistance would be less of a problem than that encountered with RT inhibitors. However, resistance was seen to accumulate rapidly (Eastman et al., 1995; Molla et al., 1996; Jacobsen et al., 1996; Markowitz et al., 1995; Condra et al., 1995; Danner et al., 1995). For example the antiviral effects of idinavir were lost within 12-24 weeks. Fortunately, increasing the dose given resulted in inhibition of virus replication and delayed drug resistance without increased toxicity (Markowitz et al., 1995; Danner et al., 1995).

An accumulation of three or more mutations was found to confer resistance to these viral variants. Also cross resistance with other protease inhibitors was evident raising concern that treatment with one protease inhibitor would confer resistance to all (Condra et al., 1995). These observations reflect the commonality in the action of these antiviral drugs and suggest that divergent, as opposed to convergent, strategies may have a more favourable outcome for combination therapy. A number of studies have been carried out to investigate the therapeutic potential of combination therapies using two and three combined drug regimes (Caliendo et al., 1994 and references within). A randomised double blind controlled trial (Delta) compared combinations of RT nucleoside inhibitors with therapeutic potential of AZT alone (Delta co-ordinating committee, 1996). Treatment with AZT combined with ddI or ddC resulted in prolonged survival when compared with AZT alone. A similar double blind trial comparing the combination of three drugs (AZT, saquinavir and zalcitabine) was found to result in a considerably more
favourable outcome than with combinations of AZT and saquinavir or AZT and zalcitabine (Collier et al., 1996). This triple combination resulted in a reduction in HIV-1 replication, increased CD4+ T cell counts and decreased the level of serum β-2-M and neopterin. Another trial investigating the combination of AZT, lamivudine and idinavir in symptomatic patients and found this triple combination resulted in the clearance of HIV viraemia to below the limit of detection by PCR (500 copies per ml; Delfraissy, 1993). More recently, the potential use of combination therapy in infants with maternally acquired HIV-1 infection was investigated (Luzuriaga et al., 1997). Previous therapeutic strategies for infants has primarily consisted of a single RT inhibitor showing moderate efficacy with limited duration. The three drug regime investigated (AZT, ddI and nevirapine) was found to reduce plasma RNA levels within four weeks of administration. Also, of the 8 infants examined 7 were found to have reduced levels of HIV-1 RNA in plasma (0.5 log) for the duration of the trial (6 months).

1.3 DIVERSITY OF HIV-1.

Following the discovery of HIV-1 it was reported that the viral genome displayed a high degree of variability (Hahn et al., 1984). In early studies restriction mapping of molecularly cloned provirus was used to describe this heterogeneity. Saag et al., described variation within three HIV-1 isolates (RJS4, WMF1 and WMF3) by restriction mapping of 17 to 27 clones (Saag et al., 1988). This study revealed 17 of 27 RJS4 clones, 9 of 17 WNF1 clones and 13 of 18
WMF3 clones had distinct cleavage patterns, although intrapatient variation was decidedly less than interpatient variation. However, two sequential isolates recovered 16 months apart (WMF1 and WMF3) contained no identical clones (estimated nucleotide divergence of 2 to 7%) indicating a rapid rate of change of HIV-1, a finding consistent with results obtained by Fisher et al (Fisher et al., 1988). Shaw et al., compared the restriction patterns of HIV-1 from a number of AIDS patients in a number of different T cell lines and found varying restriction patterns (Shaw et al., 1984). HIV-1 has been cloned and sequenced by a number of independent research groups and a large number of nucleotide differences were detected (Muesing et al., 1985; Sanchez-Pescador et al., 1985; Ratner et al., 1985). Both restriction mapping and complete sequencing of clones proved to be very labour intensive and the resolution obtained by restriction mapping was low. The introduction of the polymerase chain reaction (PCR) heralded a new era in molecular biology making it possible to amplify small quantities of DNA, such as those found in PBMC’s (Ou et al., 1988). In a study by Goodenow et al., two regions of the HIV-1 genome were amplified by PCR (gag and env segments) and multiple clones sequenced (Goodenow et al., 1989). They detected sequence heterogeneity in both regions with varying degrees of complexity. A number of other studies have been carried out which report findings of genomic heterogeneity in HIV-1 (Hahn et al., 1986; Wong-Staal et al., 1985; Alizon et al., 1986; Ratner et al., 1985; Benn et al., 1985), both within and between patients, reminiscent of other RNA viruses such as influenza A (Young et al., 1979; Webster et al., 1982), poliovirus (Nottay et al., 1981), foot and mouth disease virus (FMDV; Domingo et
al., 1980), enterovirus 70 (Takeda et al., 1984) and other retroviruses such as visna virus (Lutley et al., 1983) and equine infectious anemia virus (Montelaro et al., 1984). Genomic heterogeneity of this nature has given rise to the concept of a quasispecies commonly used to describe a population of closely related yet genetically distinct viral variants.

There is a great deal of discussion regarding the generation of diversity in HIV-1. The complex nature of the HIV-1 life cycle provides many opportunities for mutation. Retroviruses require three enzyme systems during replication, reverse transcriptase, cellular DNA polymerase and RNA polymerase, of which the most attention has centred on the reverse transcriptase. Secondly, HIV-1 has the ability to transfer from one template to another during DNA synthesis, known as strand transfers or 'jumps', which are required to generate the LTR’s at either end of the viral genome. However, other factors may enhance genetic diversity such as the high turnover rate of HIV-1 and various selective forces imposed by the host immune response and requirements for cell tropism.

Genetic variability is determined by three separate variables: the mutation rate per replication cycle, the number of replication cycles in a given time and the selective advantage or disadvantage of a particular variant. A number of studies have attempted to calculate the \textit{in vivo} mutation rate of a number of retroviruses. However, experiments of this nature are problematic in that the accumulation of mutations in a single round of replication must be measured to eliminate the effects of selection. In order to circumvent this difficulty Dougherty and Temin used a retroviral DNA vector system to measure the mutation rate of spleen necrosis virus
(SNV; Dougherty et al., 1988), using selectable marker or reporter genes, in which the replication cycle was limited to one round. They determined that the mutation rate for a single base pair substitution was $2 \times 10^{-5}$ per base pair per replication cycle and an insertion rate of $10^{-7}$ per base pair per replication cycle. Using a different SNV based vector, the base substitution mutation rate over a longer sequence was found to be $0.7 \times 10^{-5}$ per base per replication cycle, corroborating the findings of Dougherty and Temin (Pathak et al., 1990). Monk et al., determined the point mutation rate of a murine retrovirus by direct examination of RNAs from viruses isolated after a single replication cycle, reporting a mutation rate of $2 \times 10^{-5}$ bases per replication cycle (Monk et al., 1992). Studies on Rous sarcoma virus have estimated the mutation rate to be in the order of $10^{-3}$ to $10^{-4}$ (Leider et al., 1988; Coffin et al., 1980). From these studies the mutation rate for HIV-1 was previously estimated to be of the order of $10^{-4}$ per base per replication cycle (Nowak et al., 1991b; Wain-Hobson, 1993b), comparable to the error rate of purified HIV-1 RT (Gregson et al., 1994). However, a recent study by Mansky et al., determined the mutation rate of HIV-1 to be $3.4 \times 10^{-5}$ mutations per base pair per replication cycle, 20 fold less than previous error rates (Mansky et al., 1995).

The point mutation rates for retroviruses are in the range of estimates for RNA viruses in general, for example, influenza virus ($1.5 \times 10^{-5}$; Parvin et al., 1986) and poliovirus ($2 \times 10^{-6}$; Sedivy et al., 1987; Parvin et al., 1986). This suggests that the extraordinary diversity of HIV-1 is not only due to a relatively high mutation rate but may also reflect the number of replication cycles and the fixation rate of mutations.
The high level of RNA virus diversity has been attributed to error prone replicating enzymes. For example, nucleotide misincorporation in retroviruses is uncorrected as RT does not encode an exonuclease activity (Battula et al., 1976; Roberts et al., 1988), and hence any errors arising during transcription are not subject to proofreading. A number of studies have been carried out to measure the fidelity of HIV-1 RT (Bebenek et al., 1989; Roberts et al., 1988; Preston et al., 1988). Using nonsense codon reversion assays, the error rate of HIV RT was 1/1700 (Roberts et al., 1988) to 1/4000 (Preston et al., 1988), translating to approximately 5 to 10 errors per HIV-1 genome per round of replication in vivo. Error rates for HIV RT are approximately 10 fold higher when compared with avian myeloblastosis virus or murine leukaemia virus, both lacking an RT proofreading mechanism, with error rates of 1/9000 (Preston et al., 1988) to 1/17000 (Roberts et al., 1988) and 1/30000 (Roberts et al., 1988) respectively. These findings suggest that the lack of a proofreading activity alone may not completely explain the infidelity of HIV-1 RT. Indeed a number of studies have described a variety of other types of errors or rearrangements in addition to base pair substitutions which may greatly increase the infidelity of this enzyme. These include frameshifts, deletions, insertions and recombination events (Katz et al., 1990; Coffin, 1992; Bebenek et al., 1989; Temin, 1993). Furthermore, these studies have revealed that certain regions of the genome have a significantly higher error rate suggesting there may be 'hotspots' for mutation in the HIV-1 genome. For example Bebenek et al., suggested the mutation rate for env was in the region of $10^{-3}$ nucleotide substitutions per site per year while that for gag was in the region
of $10^{-4}$ nucleotide substitutions per site per year (Bebenek et al., 1989). Indeed, the env region itself is composed of conserved and hypervariable regions (Willey et al., 1986; Starcich et al., 1986; Modrow et al., 1987). Regions of hypermutation have also been documented, characterised by a large number of G to A changes particularly within GpA dinucleotides (Bebenek et al., 1989; Pathak et al., 1990; Vartanian et al., 1991). It was proposed that hypermutation may be due to slippage and realignment during reverse transcription, termed dislocation mutagenesis (Wain Hobson, 1992; Bebenek et al., 1989; Vartanian et al., 1991).

Although the above experiments reveal a high rate of mutation they do not distinguish between the enzyme systems utilized during HIV-1 replication. The fidelity of RNA polymerase II is not known but it is likely to be of the same order as other viral RNA polymerases i.e. approximately $10^{-4}$. Therefore, errors incorporated in genomic RNA during synthesis may be equally important as those produced by RT during transcription in the generation of sequence variation. An alternative mechanism has been proposed, by Temin, who suggested variation may be a direct consequence of strand transfers during DNA synthesis, required to generate LTR sequences at either end of the proviral DNA (Temin, 1993). Hence, (-) strand synthesis may transfer back and forth between the two RNA templates present, using a similar mechanism to the production of the LTR’s, increasing genetic diversity.

Genetic recombination is frequently observed during the life cycle of retroviruses. Both avian and murine type C viruses display a high frequency of genetic recombination (Wyke et al., 1979; Vogt, 1971; Linial et al., 1979; Hunter,
Recombination within lentiviruses has not been well documented although recombination between HIV-1 genomes has been shown to occur (Hu et al., 1990; Vartanian et al., 1991; Srinivasan et al., 1989; Clavel et al., 1989). The above studies may have produced somewhat erroneous results since recombination events described occurred in tissue culture systems and may not be observed during in vivo infection. Recently a study by Robertson et al., analysed a number of published HIV-1 sequences and reported a large number of apparent recombinant viruses (Robertson et al., 1995), implying co-infection with divergent HIV-1 stains may not be as rare as previously predicted (Sabino et al., 1994). Interestingly all the possible recombinant viruses originated from geographic regions where multiple subtypes are known to co-circulate, for example, Central Africa, South America and South East Asia (Clavel et al., 1989). Therefore, recombination is increasingly being recognised as a biological phenomenon among HIV-1 isolates/subtypes and may play an important role in viral evolution.

Evolution of the virus is greatly enhanced by the very high rates of replication. It has been known for some time that HIV-1 actively replicates throughout the course of infection (Embretson et al., 1993; Pantaleo et al., 1993). However, recent findings have revealed that HIV-1 infection is a highly dynamic process in which very high levels of virus replication and CD4+ cell turnover are apparent. By treating patients with a number of antiviral drugs Wei et al., and Ho et al., were able to show that both virus and CD4 cells in the peripheral blood exhibit fast rates of turnover (Ho et al., 1995; Wei et al., 1995). Extrapolating from changes seen in the blood they estimated that between $10^8$ and $10^9$ virions were
cleared daily. Since a steady state is observed during infection the number of virions cleared daily must be equal to the number produced daily. Therefore with approximately $10^9$ virions produced each day the scope for creating diversity must be extremely high.

Regardless of how the extensive variation in HIV-1 is generated the ultimate result will be determined by the selective forces acting on HIV-1 variants. For example, a number of studies have described the isolation of minor variants upon in vitro cultivation in PBMC's, suggesting the preferential selection of those viruses best adapted to tissue culture conditions (Vartanian et al., 1991; Meyerhans et al., 1989). The degree of diversity among different strains of HIV-1 may be partly due to selection by the host immune system. This is supported by in vivo studies which have demonstrated the generation of viruses that have escaped neutralization by serum antibodies (Robert-guroff et al., 1986; Reitz et al., 1988; McKeating et al., 1989). Similarly, resistant mutants can also be selected for in the presence of antiviral drugs, such as AZT (Larder et al., 1989). Differences have also been reported regarding the cellular host range among HIV-1 strains. Various studies have shown that some viruses grow in certain T, B and monocyte cell lines while others do not (Schwartz et al., 1989; Sakai et al., 1997; Levy et al., 1985; Fenyo et al., 1988; Cloyd et al., 1990). Different strains have also been found during early and late stages of infection showing substantial variability in replicative abilities in T cell lines and macrophages (Schwartz et al., 1989; Fenyo et al., 1989). Similarly, selective forces can act on coding regions which may result in the accumulation of mutations that generate new functions. For example, various
groups have shown that the sequence of the V3 envelope region can influence tropism for T cell lines or macrophages (Westervelt et al., 1992; Westervelt et al., 1991; Shioda et al., 1991; O’Brien et al., 1990; Hwang et al., 1991; Rhim et al., 1991; Chesebro et al., 1992). Therefore, diversity of HIV-1 may facilitate infection of different cell types in vivo.

1.4 TROPISM OF HIV-1.

1.4.1 HIV-1 INFECTION OF DIFFERENT CELL TYPES.

HIV infects a wide variety of tissues in vivo, and has been detected in a number of human cell lines in culture. Using various techniques, such as cell culture, in situ hybridization, immunohistochemistry, electron microscopy and PCR, HIV has been detected in vivo in cells of the haematopoietic lineage (Freedman et al., 1991; Livingstone et al., 1996; Tsubota et al., 1989; Patterson et al., 1994; Knight et al., 1993; Ali et al., 1993; Uittenbogaart et al., 1996; Shattock et al., 1996; Nicholson et al., 1986; Levy et al., 1985; Scottalgar et al., 1993; Kunzi et al., 1993; Gartner et al., 1986; Cullen, 1991), brain (Wiley et al., 1996; Wiley et al., 1986; Pang et al., 1990; Koenig et al., 1986; Price et al., 1988; Li et al., 1991; Moses et al., 1993; Wiley et al., 1990; Gartner et al., 1986), gastrointestinal tissue (Kotler, 1993; Pumarola Sune et al., 1987; Gill et al., 1992), heart (Luginbuhl et al., 1993; Grody et al., 1990), lungs (Plata et al., 1990; Dolei et al., 1992; Chayt et al., 1986), kidneys (Cohen et al., 1989), adrenals (Barboza et al., 1992), eye
(Cantril et al., 1988), salivary glands (Qureshi et al., 1995), cervix (Nuovo et al., 1993), prostate (da Silva et al., 1989), testes (Bagasra et al., 1994; da Silva et al., 1989) and skin (Tschachler et al., 1987). Similarly, cell culture studies have shown a wide variety of human cells to be susceptible to infection with HIV-1. Initial studies carried out to delineate the host cell range of HIV-1 suggested that CD4+ lymphocytes were preferentially infected (Klatzmann et al., 1984a; Coffin et al., 1986a; Dalgleish et al., 1984a).

The concept of differential tropism arises from several studies which have shown HIV-1 strains differ in their ability to productively infect various human cells in in vitro culture (Folks et al., 1986; Evans et al., 1987; Levy, 1988; Kikukawa et al., 1986; Levy et al., 1985). Established T cell lines expressing similar amounts of CD4 antigen on their surface displayed widely varying degrees of virus production and induction of cytopathic effect (Evans et al., 1987; Kikukawa et al., 1986). A similar variation in virus production was observed following infection of human peripheral blood lymphocytes, from a number of seronegative individuals, with HIV-1 (Folks et al., 1986; Evans et al., 1987).

The host range or tropism of a virus may be controlled at the entry level or post entry level. In some retroviral systems tropism of the virus is determined at the entry level by a receptor-ligand interaction at the cell surface. For example, in feline leukaemia virus (FeLV) the 3'pol-5'env region of the FeLV genome determines the host range of the virus (Reidel et al., 1988). The envelope protein of Avian leukosis virus, gp85, specifically interacts with distinct cellular receptors (Dorner et al., 1986). However, in other retroviral systems the expression of other
viral genes contributes to the host range of the virus. It has been suggested that sequences within murine leukaemia virus (MuLV) LTR region and \textit{gag} gene interact with cellular factors influencing host range and tissue tropism (Evans \textit{et al.}, 1987; Celander \textit{et al.}, 1984). In a murine retrovirus, involvement of multiple genes has been suggested to enhance tropism and oncogenesis (located mainly in the 3’ half of the genome, although sequences in the 5’ half have also been identified; Holland \textit{et al.}, 1985). When the effect of each gene was examined alone a markedly reduced effect was apparent and only moderately increased when using a combination of two.

Several studies have identified the CD4+ molecule as the major receptor for HIV-1 infection (Fisher \textit{et al.}, 1988; McDougal \textit{et al.}, 1986; McDougal \textit{et al.}, 1986; Maddon \textit{et al.}, 1986). Maddon \textit{et al.}, expressed the CD4+ molecule in a number of human CD4- cell lines and found that expression of the CD4+ molecule was sufficient to render human cells susceptible to HIV-1 infection (Maddon \textit{et al.}, 1986). McDougal \textit{et al.}, carried out binding and blocking experiments to show the direct binding of HIV-1 to CD4+ molecules and reciprocally, the inhibition of this action with anti-CD4+ antibodies (McDougal \textit{et al.}, 1986b). A number of groups have constructed truncated soluble forms of the CD4+ molecule (Berger \textit{et al.}, 1988; Deen \textit{et al.}, 1988; Smith \textit{et al.}, 1987). All were shown to block HIV-1 infection of CD4+ cells measured by the inhibition of infectivity, syncytium production and binding. Further studies have delineated the region on the CD4 molecule involved in the highly specific interaction with the virion surface proteins, namely the D1 domain (see section 1.2.8.1).
Although CD4 is recognised as the major receptor for HIV, it has been known for some time that CD4 expression alone was not sufficient for HIV infection. This was first demonstrated by Maddon et al who showed that when human CD4 is expressed in mouse cells, HIV is unable to gain entry (Maddon et al., 1986). Subsequently, a number of cells that do not express the CD4 antigen on their surface have been productively infected with HIV-1 (Li et al., 1990; Chesebro et al., 1990; Cao et al., 1990; Tateno et al., 1989; Harouse et al., 1989; Folks et al., 1988; Chiodi et al., 1987; Cheng Mayer et al., 1987; Clapham et al., 1989; Dewhurst et al., 1987). HIV molecular clones have been transfected into a wide variety of cells including human T lymphocytes and monocyte cell lines (permissive for HIV-1 infection) and human, mouse, mink and monkey fibroblast cell lines (resistant to direct HIV-1 infection; Nowak et al., 1991b; Cichutek et al., 1992b). Both these studies demonstrate the production of infectious virus particles in all cell lines examined, suggesting the block to HIV-1 infection primarily occurs at the cell surface. Intracellular mechanisms may also participate in controlling virus replication since virus production by human and animal fibroblast cell lines was greatly reduced when compared with that of human lymphocytes (Nowak et al., 1991b). Similar findings were observed using phenotypically mixed particles or pseudotypes produced through co-infection with HIV-1 and an animal retrovirus. For example, murine amphotropic retrovirus (Spector et al., 1990), or the xenotropic or dual tropic mouse type C virus (MuLV; Canivet et al., 1990). Both human and animal cell lines became productively infected, however replication was best in human cells and very limited in murine and avian cells. These findings lead
to the speculation that a co-factor or possible second receptor may be required for infection. A number of candidates have been proposed as accessory molecules, including MHC class I and II (Grassi et al., 1991), LFA-I (Hildreth et al., 1989), tryptase II (Hattori et al., 1989), CD26 (Alizon et al., 1994; Callebaut et al., 1993) and a number of undefined molecules (Henderson et al., 1993). Recently a number of proteins from the superfamily G-protein-coupled-receptors with a characteristic seven transmembrane domain have been implicated as co-factors required for HIV-1 entry. The first to be characterised was CXCR-4 which facilitates the infection of T cell lines but not infection of macrophage tropic stains of HIV-1 (Feng et al., 1996). A second co-receptor for HIV which facilitates infection of macrophage tropic strains was identified simultaneously by three research groups and is called CCR-5 (Alkhatib et al., 1996; Dragic et al., 1996; Deng et al., 1996).

1.4.2 DIFFERENTIAL PROPERTIES IN THE HOST RANGE OF HIV-1.

A large number of studies have been carried out to investigate the biological properties of variants of HIV-1 infecting different tissues. Several studies have demonstrated that different HIV-1 isolates infect monocytes more readily than T cells (Koyanagi et al., 1987; Gartner et al., 1986; Cheng Mayer et al., 1990; Gendelman et al., 1990; Schuitemaker et al., 1991; Briesen et al., 1990; Valentin et al., 1990). Gartner et al., compared viral isolates obtained from lung and brain derived macrophages, and found they displayed a significantly higher ability to infect macrophages than T cells (Gartner et al., 1986b). These findings are in
contrast with the properties of the laboratory adapted isolate, HIV-IIIb, which showed a 10,000 fold lower ability to infect macrophages than T cells. A second study compared isolates from cerebrospinal fluid (CSF) and brain tissue from patients presenting with AIDS encephalopathy (Koyanagi et al., 1987). Both isolates were found to replicate efficiently in PBL’s, although efficient and productive infection of macrophages was only accomplished using isolates obtained from brain tissue.

Kinetic studies have described differences in cytopathogenicity and replicating abilities among HIV-1 isolates. Initially pathogenic strains of HIV were identified using the MT-4 plaque assay which identified HIV variants with enhanced replication kinetics and cytopathogenicity isolated from patients with clinical disease compared to healthy seropositive individuals (Tateno et al., 1988). Subsequently, an increasing number of studies have reported a relationship between the clinical severity of HIV-1 infection and the in vitro replication potential of HIV-1 (Tersmette et al., 1988; Fenyo et al., 1989; Sakai et al., 1988; Fenyo et al., 1988; Schwartz et al., 1989; Asjo et al., 1986). It has been suggested that HIV-1 exists in two different states early and late in infection. Viruses isolated from asymptomatic individuals are thought to be less virulent, displaying low replication rates, do not produce syncytia and replicate efficiently in macrophages but not transformed T cell lines. These have been designated as slow/low, non syncytium inducing viral isolates (NSI). During disease progression more virulent strains can be isolated displaying higher replication rates and decreased ability to replicate in macrophages (SI). In approximately 50% of all infected individuals SI isolates
emerge which have the capacity to cause syncytia and replicate in transformed T cell lines. These have been designated rapid/high viral isolates.

In order to understand the differences involved during replication of these isolates Schwartz et al., were able to distinguish between slow/low and rapid/high isolates using the CAT bioassay (Schwartz et al., 1989). Rapid/high isolates can activate CAT expression in T lymphoid and monocytoid cell lines, whereas slow/low viruses only activate CAT expression in monocytoid cell lines. This assay is based on the ability of HIV-1 isolates to activate an LTR derived from the HXB2c molecular clone (isolated from the laboratory adapted strain HIV_{IIIIB}). A block in productive infection of T cell lines by the slow/low isolates must therefore take place early in the infection cycle and is probably associated with a lack of entry.

1.4.3 ROLE OF THE V3 REGION OF gp120 IN CELLULAR TROPISM

Regions located in the env gene of gp120 outwith the CD4 binding domain (carboxy terminal) have been associated with viral tropism. These include the V1, V2 and V3 hypervariable regions. While most HIV-1 stains are capable of replication in primary lymphocytes, replication in other cell types is more restricted (Schuitemaker et al., 1991). HIV-1 isolates capable of infecting macrophages are generally unable to infect T cell lines and vice versa (Cheng-Mayer et al., 1988). The restriction of HIV-1 isolates for replication in macrophages is thought to be at the level of virus entry, the critical determinant being within the V3 region of
gp120. V3 also serves as a major target for neutralizing antibodies (Rusche et al., 1988; Matsushita et al., 1988; Javaherian et al., 1989). Additional studies have shown that V3 contains epitopes that elicit CTL responses (Takahashi et al., 1989; Clerici et al., 1991). Extensive sequence analysis has been carried out on the V3 region to identify the regions of the virus responsible for macrophage tropism. Hwang et al., demonstrated that the V3 loop was a major determinant of cell tropism. They substituted a 20 amino acid sequence from the V3 loop of a macrophage tropic isolate (HIV_{Bal}) into a T cell tropic virus (HTLV_{IIIB}). This provirus (IIIB/V3-BaL) replicated in macrophages equally as well as the HIV_{Bal} proviral clone but was no longer able to infect T cell lines (Hwang et al., 1991). Shioda et al., constructed recombinant viruses between molecular clones of HIV-1 isolates showing differential abilities to infect T cell lines and macrophages (HIV-1_{SF2}: T cell line tropic and HIV-1_{SF162}: macrophage tropic; Shioda et al., 1991a). They found that infection of primary macrophages was associated with a 159 amino acid region encompassing the V3 hypervariable region. A similar finding was reported by O’Brien et al., following examination of the molecular clone HIV-1_{JR-FL} (O’Brien et al., 1990).

A number of studies have reported an accumulation of amino acid substitutions within the V3 region during the course of infection, some of which alter the cellular host range of HIV-1 variants (Morris et al., 1994; Shioda et al., 1994; Wolfs et al., 1991; Chesebro et al., 1992). Macrophage-tropic isolates showed remarkable similarity in V3 while T cell tropic isolates were found to be highly heterogenous (Chesebro et al., 1992; Yamashita et al., 1994). Morris et al.,
found that a single amino acid substitution at position 312 (threonine to alanine) in the V3 loop resulted in the generation of a revertant virus, infectious for Sup-T1 cells (T cell-line) but not for AA5 cells (Epstein barr virus transformed B cell line highly permissive for HIV-1 infection). A subsequent mutation at position 306 (arginine to serine) restored the ability of the virus to infect AA5 cells (Morris et al., 1994). Substitution of an acidic amino acid (aspartic acid) for a basic amino acid (lysine) at position 30 (from the first cysteine) in the V3 loop altered the capabilities of a recombinant virus to grow in a human T cell line (Oka et al., 1994; Shioda et al., 1994). A study by Wolfs et al., observed a reduction in the binding affinity of a patients antibodies following an amino acid substitution at position 308 in the V3 loop. Subsequently, a specific humoral immune response was mounted against this new variant indicating antigenic variation within the viral population (Wolfs et al., 1991). This amino acid has previously been reported to be a critical determinant of antibody binding specificity (McKeating et al., 1989).

Analysis of V3 loop sequences obtained from cloned and primary HIV-1 isolates has revealed a distinct pattern of amino acid substitutions within this region that correlated with virus phenotype (Millich et al., 1993). It was found that a combination of non-conservative basic amino acid substitutions in positions 11, 24, and 32 plus a basic or uncharged amino acid residue at position 25 were predictive of an SI phenotype. The presence of an acidic amino acid at position 25 was found to correlate with an NSI phenotype. These findings are consistent with previous analyses of this hypervariable loop (de Jong et al., 1992; Koup et al., 1994; Fouchier et al., 1992; Shioda et al., 1992; Shioda et al., 1994; Morris et al., 1994).
Several studies have shown that amino acid substitutions flanking the GPGR sequence (found predominantly in V3 sequences) were sufficient to alter virus tropism to T cell-line tropic (de Jong et al., 1992; Chesebro et al., 1992). De Jong et al., found a correlation between a high charge in the V3 region, due to basic amino acid substitutions, and an SI phenotype (de Jong et al., 1992). A more recent study has examined the V3 loop in a number of HIV-1 group O isolates of known phenotype, as determined by the MT-2 assay (de Jong et al., 1996). SI isolates were found to have a high net positive charge with a positive amino acid at positions 11 or 25, while NSI isolates had a low positive net charge accompanied by a positive charge at position 37.

A recent study carried out by Donaldson et al., investigated the in vivo distribution and cytopathology of HIV-1 (Donaldson et al., 1994a). The aim of this study was to investigate whether there were consistent differences between HIV-1 variants obtained from lymphoid tissue (lymph node, spleen and PBMCs) with those from non-lymphoid tissues (brain, spinal cord, lung and colon). Comparisons of variants from lymphoid and non-lymphoid tissues between asymptomatic and symptomatic individuals were not possible as it has previously been shown that HIV-1 is not detected in non-lymphoid tissue until the onset of AIDS (Donaldson et al., 1994b). In the group of patients dying in AIDS, distinct HIV-1 populations were detected in different organs, although a general observation was of a common set of sequences present at varying frequencies within the tissues sampled. For example, the major V3 sequence present in brain and spinal cord (12 of 17 and 6 of 7 respectively) was also detected in this individual's colon, lung and lymph node.
as minor variants (5 of 16, 7 of 15 and 1 of 16 respectively). In some instances, however variants appeared to be more restricted in distribution. For example, variants detected in the colon of patient 4, the brain of patient 5 and the colon of patient 6 were not detected elsewhere in each respective patient (10 of 16, 17 of 17 and 12 of 15 respectively). In addition, in patient 9 two distinct populations were detected in left and right brain tissue suggesting a random component in the in vivo distribution of env sequences. Previously, Millich et al., proposed a method whereby the phenotype of a virus could be inferred from the overall charge of the V3 sequence combined with the number of observed differences from a subtype B consensus sequence. This type of analysis provided an almost complete spatial separation of published HIV-1 variants of known phenotype (Millich et al., 1993). This method was employed to infer the phenotype of post mortem sequences in this study. Almost all sequences were found to have a predicted NSI/macrophage tropic phenotype with low charge and few changes from the subtype B consensus irrespective of disease status.

1.4.4 NEUROTROPISM OF HIV-1.

Virological analysis, of HIV-1 infection in the CNS, has shown infection of the brain occurs preferentially in microglia (Koyanagi et al., 1987; Vazeux et al., 1987; Bagasra et al., 1996; Koenig et al., 1986; Khabbaz et al., 1994). In addition, only macrophage tropic HIV-1 isolates establish productive infection in primary human brain explant cultures (Watkins et al., 1990). T cell tropic strains were
generally found unable to replicate in brain explant cultures (Watkins et al., 1990). These findings lead to the suggestion that specific neurotropic strains of HIV-1 may exist. Previously, Power et al., reported that a histidine at position 305 in the V3 loop was predominant in patients with dementia, suggesting that this residue may determine neurotropism (Power et al., 1994). However, later studies have been unable to reproduce these findings. Di Stephano et al., (1996) analyzed the V3 region from brain and CSF of individuals at different stages of disease and found a histidine at position 305 in both demented and nondemented individuals. A similar study reported no evidence for conserved differences in this region between brain and spleen isolates (Reddy et al., 1996). A comparison of isolates derived from blood and brain tissue in the V3 region described a monophyletic population of brain derived compared to blood derived isolates (Korber et al., 1994). In contrast, two further studies comparing isolates from the CSF and blood observed no consistent differences between nucleic acid sequences from the distinct compartments (Kuiken et al., 1995; Keys et al., 1993). A more recent study examined the tissue distribution of HIV-1 variants from HIV-1 infected individuals dying in AIDS and those dying from unrelated causes while asymptomatic. Although organ-specific differences in V3 were observed, the majority of individuals were found to harbour NSI macrophage tropic isolates regardless of tissue origin or disease status (Donaldson et al., 1994a). Infection of the CNS may be determined solely by the ability to infect cells of the macrophage lineage, as conclusive evidence for a specific neurotropic HIV-1 variant has not been obtained. The limited amino acid substitutions observed in the V3 region, between different
tissues, would be unlikely to alter biological properties to such an extent as to confer different tropic properties.

1.5 PATHOGENESIS OF HIV-1 INFECTION.

Since the identification of HIV-1 as the aetiological agent for AIDS enormous progress has been made in our understanding of this virus. As discussed in sections 1.2.2 to 1.2.8 the biological and molecular characterization of HIV has revealed a complex genomic structure that encodes proteins with structural and regulatory functions. It is also clear that an important event in the pathogenesis of HIV is the binding of the HIV surface glycoprotein, gp120, to the CD4 receptor present on the surface of CD4+ lymphocytes and a number of other cells (see section 1.2.8.1). However, the precise mechanisms whereby HIV causes the dramatic decline in CD4+ T cells, resulting in severe immunosuppression and the development of opportunistic infections and neoplasms, are presently unknown or speculative at best. A number of mechanisms have been proposed involving both direct and indirect methods by which HIV could destroy CD4+ T cells. Similarly, the way in which this increase in immunosuppression relates to disease progression is currently a topic of great discussion. Why does the immune system fail to control HIV-1 infection, ultimately resulting in the plethora of clinical manifestations associated with AIDS? Again a number of mechanisms have been proposed involving the direct action of HIV and indirect mechanisms related to the host immune response.
1.5.1 MECHANISMS OF HIV-1 INDUCED CYTOPATHICITY.

1.5.1.1 DIRECT MECHANISMS.

The direct relationship between the high levels of viraemia and a precipitous drop in CD4+ T cells during the acute phase of HIV-1 infection suggests that HIV-1 is directly responsible for the death of CD4+ cells. The mechanisms involved, however, are not fully understood. A number of studies have suggested disturbances in membrane permeability may lead to cell death. Formation of holes in the cell membrane may occur as a result of virion budding. In a cell that is producing very high levels of HIV-1 virions, consistent budding from the cell surface may weaken the integrity of the cell membrane resulting in cell lysis. Cloyd and collaborators reported findings which demonstrated a perturbation of the host cell membrane and lipid synthesis leading to an inability of the cell to control the influx of Ca²⁺ which accumulated in the cell (Lynn et al., 1988; Cloyd et al., 1991). The loss in control of the intracellular ionic strength may lead to the impairment of cellular functions and the eventual lysis of the cell.

Previously, it has been shown that HIV infected individuals, regardless of clinical stage, exhibit persistent plasma viraemia in the range of $10^2$ to $10^7$ virions per ml (Piatak et al., 1993). More recently a number of novel experiments have been carried out to measure the dynamics of HIV-1 replication in vivo. Wei et al. (1995) and Ho et al. (1995), recently provided estimates of viral replication and CD4+ T cell destruction in the peripheral blood following the administration of
potent anti-retroviral drugs. Both studies showed very high rates of turnover, with approximately 30% of the total virus population and 5% of the total CD4+ T cell population being replaced each day. By measuring the reduction in plasma viraemia following antiretroviral treatment both groups estimated the half life of the virion population to be approximately two days. The rate of virus production was calculated by factoring in the patients blood volume and initial virus density and was found to be approximately $10^9$ virions per day consistent with later studies (Perelson et al., 1996). Therefore, the vast turnover of CD4+ cells throughout the course of infection may have a significant effect on the pathogenesis and clinical course of HIV-1 infection.

HIV-1 has been found to persist in substantial amounts in cells as unintegrated linear DNA (Shaw et al., 1984). Pauza et al.(1990), reported that the accumulation of unintegrated DNA was directly associated with death of the culture. A similar phenomenon has been reported for a number of animal retroviruses where accumulation of unintegrated DNA in the cytoplasm of infected cells has been associated with subsequent cell death (Keshet et al., 1979; Weller et al., 1980). These studies support the conclusion that high levels of unintegrated viral DNA may be toxic to the cell and could contribute to the observed depletion of CD4+ cells. However, in a more recent study, single cell lysis was reported to occur in the absence of the accumulation of unintegrated DNA (Bergeron et al., 1992), suggesting other mechanisms are involved in cell death.

High levels of viral RNA have been detected in the cytoplasm of HIV-1 infected cells in vitro (Koga et al., 1988; Somasundaran et al., 1988).
Somasundaran et al. (1988), reported HIV-expressing cells contained from 300,000 to 2,500,000 copies of viral RNA per cell. However, more recent work has indicated substantially lower copy numbers of viral RNA, in the region of 40,000 copies per cell (Robinson et al., 1990). Koga et al., have shown that HIV infection of a T cell line (HUT-78) results in the accumulation of high levels of large molecular weight heterodisperse RNAs (containing repetitive sequences; Koga et al., 1988). More specifically, significantly increased levels of viral RNA in mitochondria compared with that in the cytoplasm or nucleus have been detected (Somasundaran et al., 1994). Therefore, the presence of high levels of viral RNA may result in the interruption of normal cellular RNA synthesis and processing, which in turn may contribute to cell dysfunction and death.

1.5.1.2 INDIRECT MECHANISMS.

A major biologic feature of HIV-1 infection is the formation of multinucleated cells in culture, formed from the fusion of uninfected CD4+ T cells with infected cells. HIV-1 envelope glycoproteins are inserted into the host cell membrane during virion assembly prior to the acquisition of the envelope as the virion buds from the cell. As a result a number of infected cells have envelope glycoproteins on their surface which are able to bind CD4 molecules on the surface of uninfected cells. In vitro studies have shown the formation of multinucleated giant cells in infected cultures which have formed through cell fusion (Yoffe et al., 1987; Lifson et al., 1986a; Lifson et al., 1986b). These giant cells died shortly after
they were formed suggesting cell fusion leading to cell death may be one mechanism to account for the progressive depletion of CD4+ cells observed in individuals infected with HIV-1 (Yoffe et al., 1987; Lifson et al., 1986a; Lifson et al., 1986b). However, a number of other studies have revealed a discordance between syncytium formation and cytopathicity reporting that HIV induced cell killing does not always involve cell fusion (Somasundaran et al., 1987; Stevenson et al., 1990). Indeed, Dedera et al., suggested that both single-cell lysis and syncytium formation occur simultaneously during acute infection (Dedera et al., 1991).

It has also been proposed that disease progression is driven by factors of the hosts immune system. When a virus enters the body a multifaceted response is mounted involving a number of cells, such as macrophages, T helper cells and natural killer cells, culminating in the production of specific antibodies and cytotoxic T lymphocytes. It has been suggested that HIV-1 may be non-cytopathic and that disease manifestations observed during infection may reflect immunopathological consequences of an anti-HIV CD8+ T cell response (Zinkernagel et al., 1991; Plata et al., 1987; Zinkernagel et al., 1994; Zinkernagel, 1995), shown to be produced throughout infection (Pinto et al., 1995; Walker et al., 1987). A parallel between HIV-1 infection and persistent infections with other viruses have been suggested (Zinkernagel et al., 1994; Mosier et al., 1993; Leist et al., 1988; Odermatt et al., 1991) where the balance between spread of the virus and the immune response is critical to the disease outcome. For example, infection of mice with a low dose of the non-cytopathic lymphocytic choriomeningitis virus
(LCMV) results in clearance of the virus and little, if any, tissue damage due to an effective CTL response. On the other hand, if LCMV spreads rapidly before an active immune response is elicited, widespread tissue damage may occur as a result of the ensuing CTL response directed against infected cells (Odermatt et al., 1991; Leist et al., 1988; Zinkernagel et al., 1994). Evidence suggests that destruction of macrophages and follicular dendritic cells may be caused by an antiviral CD8+ response (Odermatt et al., 1991; Leist et al., 1988). Similarities between LCMV infection and HIV infection have been suggested since both show tropism for different cell types, cause severe immunosuppression, and to some extent, establish persistent infections (Odermatt et al., 1991; Leist et al., 1988). Walker et al. proposed that a similar mechanism of destruction may operate during HIV infection (Walker et al., 1986). However, a number of studies have proposed a protective role for CTL’s in HIV-1 infection (Sharpless et al., 1992; Mackewicz et al., 1991; Wolinsky et al., 1996). A recent study demonstrated that there was an inverse relationship between progression to disease with the humoral immune response or with the CTL response (Wolinsky et al., 1996). They found an increased CTL activity in those individuals showing a slower rate of disease progression, suggesting CTL activity may protect the host.

Various autoimmune mechanisms may inadvertently kill uninfected cells as 'innocent bystanders'. Uninfected CD4+ T cells may bind free gp120 molecules to their surface CD4 receptor giving them the appearance of an infected cell and therefore become a target for lysis by antibody-dependent cellular cytotoxicity (ADCC) responses (Weinhold et al., 1989; Lyerly et al., 1987). HIV-specific CTL

106
may also mistakenly kill uninfected CD4+ cells that have captured and processed soluble gp120 produced by infected cells, presenting it on their surface (MHC class II restricted antigen presenting cells; Lanzavecchia et al., 1988; Siliciano et al., 1988). In this regard, CTL that can lyse uninfected CD4+ T cells have been reported in HIV infected individuals but not HIV infected chimpanzees, suggesting the lack of CTL response in chimpanzees may account for the lack of progression to AIDS (Zarling et al., 1990). It has also been shown that natural killer cells can mediate ADCC against uninfected cells that have HIV antigens on their surface (Katz et al., 1988).

More recently, superantigens have been implicated in HIV-mediated immunosuppression. Superantigens differ from typical antigens in their interactions with T cells. Typical antigens must bind in the groove of the MHC class II molecule interacting with both α and β subunits of the T cell receptor (TCR). Superantigens, on the other hand, are presented on the outside of the MHC class II molecule and only require to bind to the Vβ subunit of the TCR. As a result superantigens can stimulate a much higher proportion of T cells than conventional antigens, followed by depletion and anergy (Kappler et al., 1980). However, it is likely that superantigens are not directly responsible for the deletion of CD4+ T cells since several studies have observed distinct Vβ perturbations in HIV infected individuals (Dalgleish et al., 1992; Imberti et al., 1991). It is therefore more likely that superantigens activate large numbers of T cells, resulting in enhanced HIV-1 infection, replication and subsequent cytopathicity. Indeed, in mice, it has been shown that murine retrovirally encoded superantigen can stimulate the proliferation
and expansion of Vβ-bearing subsets of T cells (Hugin et al., 1991).

Inappropriate induction of apoptosis has been proposed as a mechanism to explain CD4+ T cell loss during HIV-1 infection. During normal thymic development, autoreactive T cells are eliminated by the induction of apoptosis, a physiological process that is a form of programmed cell death (Jenkinson et al., 1989; McConkey et al., 1989; Smith et al., 1989; Duvall et al., 1986). Apoptosis is characterized by an increase in cytoplasmic calcium levels and the function of a Ca$^{2+}$-dependent endogenous endonuclease that produces fragmentation of nuclear DNA (Sasaki et al., 1996; Wylie et al., 1991). Apoptosis in HIV-mediated pathogenesis is supported by several studies. It has been shown to occur in vitro after HIV-1 infection of MT-2 lymphoblasts or activated PBMC (Laurent-crawford et al., 1991; Terai et al., 1991). Stimulation of CD4+ T cells with MHC class II dependent superantigens or pokeweed mitogen resulted in the cell death of CD4+ cells from asymptomatic individuals. However, activation-induced cells death did not occur in CD4+ T cells from seronegative donors (Groux et al., 1992). Also, large amounts of histones and fragmented DNA have been detected in the nucleoplasm of infected cells by SDS PAGE, immunoblot assays and Coomassie blue staining (Laurent-crawford et al., 1991). Similarly, autoantibodies to histones and DNA have been found to circulate in the serum of infected patients consistent with the release of nuclear antigens by apoptosis (Muller et al., 1992). The mechanism by which induction of apoptosis occurs during HIV-1 infection remains unclear. A number of studies have suggested an initial interaction involving the cross-linking of CD4 and gp120 followed by T cell activation, inducing apoptosis (Ameisen et al., 1991;
These studies demonstrate a central role for env-CD4 interactions in initiating apoptosis however the subsequent mechanisms involved have not been elucidated. Various mechanisms have been proposed which include the direct action of HIV, HIV viral proteins, CD4 antibodies, gp120-antibody complexes, cytokines and superantigens.

A number in vitro studies have reported the regulation of HIV expression by cytokines. Various cytokines, such as tumor necrosis factor (TNF) -α, interleukin (IL) -6, granulocyte-macrophage colony stimulating factor (GM-CSF), interferon (INF) -γ, transforming growth factor (TGF)-β, IL-6 and IL-3, have been shown to induce HIV expression in vitro (Poli et al., 1992; Hober et al., 1993; Poli et al., 1990a; Butera et al., 1993; Poli et al., 1990b; Matsuyama et al., 1989). Several studies have suggested that HIV-1 itself can upregulate cytokine expression. For example, increased production of TNF-α and IL-6 have been detected in HIV-infected cells in vitro (Molina et al., 1990). Similarly, increased expression of TNF-α and IL-6 are secreted from PBMC from AIDS patients (Brix et al., 1990; Wright et al., 1988). Therefore, abnormal expression of cytokines may act to upregulate HIV-1 expression and hence may contribute to the observed pathogenesis. Indeed, TNF-α preferentially induces lysis of HIV-infected cells suggesting that it may also function in the destruction of CD4+ T cells (Hober et al., 1992).

1.5.2 NEUROPATHOGENESIS OF HIV-1 INFECTION.
Infection with HIV-1 is frequently complicated in the later stages by the AIDS dementia complex (ADC), a neurological syndrome characterized by abnormalities in cognition, motor performance and behaviour. Exactly how HIV-1 enters the brain is uncertain. Infected brain macrophages may originate from expansion of peripherally infected monocytes that carry HIV-1 into the brain ('Trojan horse' hypothesis), in a similar manner to that proposed for visna virus (Peluso *et al.*, 1985). Alternatively, virus may cross the blood-brain barrier as cell free virus or via trafficking by infected T cells (Michaels *et al.*, 1988). Recent studies have suggested that infection of brain microvascular endothelial cells (BMVECs) may be one of the primary routes whereby HIV-1 gains entry into the CNS (Nottet *et al.*, 1995; Bagasra *et al.*, 1996). Indeed, infection of BMVECs has been demonstrated in cell culture (Moses *et al.*, 1993), and *in vivo* for SIV (Mankowski *et al.*, 1994). Another possible route of infection may be the productive infection of the choroid plexus via the penetration of cell-free virions from the CSF. Bagasra *et al.*, recently reported productive infection (by detection of mRNA) of the choroid plexus of AIDS patients by dual *in situ* PCR (Bagasra *et al.*, 1996). Studies to date suggest HIV-1 is predominantly localized within blood-derived brain macrophages, microglia and multinucleated giant cells (MGC; Vazeux *et al.*, 1987; Koenig *et al.*, 1986; Wiley *et al.*, 1986; Bagasra *et al.*, 1996). There are few reports of HIV-1 infection of neurons, oligodendrocytes and astrocytes and it is thought these cells are rarely infected *in vivo*, if at all (Gyorkey *et al.*, 1987; Wiley *et al.*, 1986). As with infection of the CNS, the mechanisms underlying the pathogenesis of HIV-1 infection in the brain are far from clear.
Several studies have detected large amounts of proviral and unintegrated HIV-1 DNA in brain tissue (Shaw et al., 1985; Vazeux et al., 1992; Pang et al., 1990). Pang et al., demonstrated a correlation between HIV encephalitis and the presence of unintegrated DNA in brain tissue similar to that observed from infection with avian leukaemia viruses (ALV) and some strains of feline leukaemia viruses (FeLV; Weller et al., 1981; Mullins et al., 1986). However, infection in brain tissue is restricted to macrophage and microglia cells and destruction of these cells would not be expected to cause the neuropathology observed in brain tissue of HIV-1 infected individuals. Similarly, the number and distribution of these infected cells does not correlate with the degree of tissue pathology observed.

Several studies have suggested that HIV-1 may effect CNS function without direct infection of neuroectodermal cells themselves. One such mechanism concerns the toxic effects of gp120 on nerve cells in vitro related to increases in intracellular calcium. Dryer et al., reported that HIV-1 env gp120 was toxic to primary rodent neurons and that cell death occurred subsequent to excessive calcium influx into the neurons (Dreyer et al., 1990). This effect was abrogated by adding calcium channel blockers. Subsequent studies suggested that the intracellular increase in Ca\(^{2+}\) and consequential neuronal injury was mediated by the combined action of gp120 with glutamate via the N-methyl-D-aspartate (NMDA) receptor (Pittaluga et al., 1996; Lipton, 1992). Noradrenaline has been implicated in learning and memory and activation of NMDA receptors is a primary event in cognitive processes in the hippocampus and cerebral cortex (Collingridge et al., 1987; McGaugh, 1989). Hence, activation of NMDA receptors may facilitate persistent
and abnormal levels of noradrenaline leading to imbalances in the processes underlying learning and memory which may result in the cognitive defects characteristic of AIDS dementia. The HIV-1 trans-activating regulatory protein, tat, has also been shown to be neurotoxic to glioma, neuroblastoma and cultured human fetal brain cells in vitro and to rats and mice in vivo (Nath et al., 1996; Sabatier et al., 1991). The active site is thought to be located between amino acids 31 and 61, the action of which alters cell permeability resulting in increased levels of intracellular Ca\(^{2+}\) and neuronal damage.

Other investigators have suggested indirect effects related to toxic substances secreted from HIV-1 infected macrophages. Cellular factors released from HIV-1 infected monocytes were found to be neurotoxic to chick and rat neurons (Giulian et al., 1990), and human brain aggregates cultured in vitro (Pulliam et al., 1991). These studies suggest HIV-1 infected macrophages, and possibly microglia, may release toxins that disrupt neurological function. A more recent study failed to confirm these findings (Benton et al., 1992), although further studies reported the requirement for cell-cell contact between HIV-infected monocytes and astroglia prior to the generation of neurotoxins (Geruś et al., 1992; Tardieu et al., 1992). The presence of activated macrophages and T cells in the CNS of HIV-1 infected individuals suggests that a number of cytokines may participate in neuropathogenesis (Tayor et al., 1992). Production of IL-1 and TNF-\(\alpha\) have been reported both in in vivo and in vitro studies. IL-1 has many activities including inducing T cell stimulation, activation of macrophages and endothelial cells, induction of fever and inflammation and stimulation of astrocytosis in vivo.
which is frequently observed in HIV encephalitis (Dinarello et al., 1987; Movat, 1987). TNF-\(\alpha\) has also been reported to be produced by macrophages and microglia in culture (Rosenberg et al., 1990; Movat, 1987), is toxic to oligodendrocytes in culture (Selmaj et al., 1990), and has been shown to be increased in the CSF of HIV-infected individuals presenting with neurological disorders. Similarly, higher levels of TNF-\(\alpha\) have been detected in symptomatic individuals compared with asymptomatic individuals (Wright et al., 1988). Other cytokines such as IFN-\(\gamma\) and IL-6 have also been implicated in CNS damage. Hence, T cells, macrophages, microglia, endothelial cells and astrocytes appear to be involved in the response to and generation of cytokines. These cytokines may modulate the immune response and may also have toxic effects in the CNS.

1.5.3 HIV-1 PATHOGENICITY AND DISEASE PROGRESSION.

An important advance in our understanding of the pathogenesis of HIV-1 infection was the observation that HIV-1 actively replicates in lymphoid tissue throughout the course of disease. Studies on the viral burden and levels of virus replication in the peripheral blood and lymphoid organs have found infection of lymph nodes to be approximately 10 times that of the peripheral blood and that active replication occurred in lymphoid tissues during the asymptomatic phase (Embretson et al., 1993; Pantaleo et al., 1993). Using in situ hybridization a number of investigations have shown restriction of viral particles to the germinal centres (Biberfeld et al., 1986; Schuurman et al., 1988; Pantaleo et al., 1993),
confirming previous reports of extracellular association of HIV with follicular dendritic cells (Armstrong et al., 1948; Baroni et al., 1986). This association may be important in infection of lymphocytes as they migrate through the lymph nodes. It has also been suggested that as disease progresses the integrity of the lymph node breaks down reducing the trapping of virus in germinal centres, allowing proliferation into peripheral sites of the body.

Recent findings have shown a rapid turnover of virus and CD4+ lymphocytes in the peripheral circulation throughout infection with HIV-1 (Ho et al., 1995; Wei et al., 1995). Both groups were also able to show the rapid replacement of wild type virus by drug-resistant virus after 14 to 28 days of drug therapy indicating the enormous capacity of HIV-1 to evolve in response to selection pressure such as that exerted by the immune system. These new findings have a number of implications for the pathogenesis of HIV-1 infection. First, the observed rates of viral and CD4+ T cell turnover were much higher than expected, indicating a dynamic process involving continuous viral infection and replication and rapid cell turnover. Second, Wei et al. (1995), also estimated the half life of PBMC by measuring the time required for the resistant virus to spread into this population and obtained a half life of 50-100 days. This substantially longer lifespan of latently infected cells suggests that virus production itself is paramount in the observed depletion of CD4+ T cells. Finally, the rapid replacement of wild type virus by mutants demonstrated the ability of the virus to evolve in response to selection pressure supporting the antigenic variation model for the subsequent progression to AIDS (see below).
Nowak and colleagues proposed the 'antigenic diversity threshold' theory to explain the breakdown of the immune system during HIV-1 infection (Nowak et al., 1991a). The central theme of this theory is that antigenic diversity is the cause and not the consequence of the observed immune deficiency in HIV-1 infection. It was proposed that HIV is capable of infecting all CD4+ cells but that the immunological response directed against the virus involves a specific response to individual strains in conjunction with a cross-reactive response to all strains (Nowak et al., 1991a). Hence, as new antigenic variants appear (escape mutants), they are able to evade elimination by the immune system. Eventually, antigenic diversity is thought to increase beyond a threshold value and exceeds the capacity of the immune response to regulate viral population growth. This model has since been adapted to include the CTL response against multiple epitopes (Nowak et al., 1995). This new improved model suggests that an antigenically homogenous population will elicit a response against a single epitope (immunodominance), while a heterogeneous population may stimulate a response against multiple epitopes which may result in a steady state number of CD4+ T cells and viral load. However, antigenic variation in the immunodominant epitope may direct immune responses to weaker epitopes reducing the immunological control of the viral population (Nowak et al., 1995).

Wolinsky et al. (1996), have studied the evolution of HIV-1 in 6 infected patients with variable rates of disease progression and refute the 'antigenic diversity threshold' theory of Nowak and colleagues. In this study two patients with high levels of plasma RNA and rapid CD4+ T cell loss maintained a
relatively homogenous population throughout the entire course of infection. In contrast, two further patients with moderate or relatively stable CD4+ T cell counts showed the highest levels of virus diversity. Therefore, no evidence was found to suggest an increase in antigenic diversity as the driving force behind the rate of CD4+ T cell loss. Interestingly, in this patient group the accumulation rate of nonsynonymous substitutions ($d_n$) was found to be higher in those patients with moderate or relatively stable CD4+ T cell counts compared with the two patients with rapid CD4+ T cell loss. The rate of synonymous substitutions ($d_s$) was similar in both groups. However, proviral sequences were obtained from PBMCs which does not distinguish between active infection and residual provirus in the memory cell population (Simmonds et al., 1991). $d_n/d_s$ ratios for those patients with moderate or stable CD4+ T cell counts were far higher (2.5 to 3.75) compared to the two patients with rapid CD4+ T cell loss (1.25 and 0.75) where a value greater than 1 is indicative of positive selection for change. Therefore, in the slower progressing patients positive selection for change was associated with prolonged rather than shortened survival. In this respect, Lukashov et al., recently reported a correlation between the accumulation of nonsynonymous substitutions and the length of the immunocompetent period from the analysis of RNA obtained from the sera of a number of progressors and nonprogressors (Lukashov et al., 1995). Wolinsky et al., also found increased CTL activity in patients with a slower rate of disease progression and very poor CTL responses in rapid progressors suggesting a protective role for the CTL response (Wolinsky et al., 1996), contrary to predictions that immunopathogenic effects of CTL are responsible for the depletion
of CD4+ T cells. A similar relationship has been reported by Harrer et al., (Harrer et al., 1994).

1.6 MOLECULAR EVOLUTION.

Molecular evolution encompasses the study of the rates and patterns of change occurring in genetic material (for example DNA sequences) and the product of such coding regions (proteins) through evolutionary time and the mechanisms responsible for these changes. This field of study also involves the reconstruction of the evolutionary history of genes and organisms, also known as molecular phylogeny, inferred from molecular data.

1.6.1 DISTANCE ESTIMATION.

Increasingly the data becoming available for the analysis of evolutionary relationships consists of nucleotide sequences. A basic process in the evolution of nucleotide sequences is the change in nucleotides with time. However, as the process of nucleotide substitution and fixation within a population is a comparatively slow process evolutionary changes within a DNA sequence can be detected through comparative methods such as distance estimation. This requires that a given sequence be compared with another sequence with shared common ancestry. The evolutionary distance between a pair of sequences is usually measured by the number of nucleotide or amino acid substitutions between them.
(pairwise distance). There are many methods for measuring evolutionary distances, although I have only discussed two such methods that are relatively simple and frequently used (the first of which, Jukes-Cantor, was used in this study: Gojobori et al, 1990; Nei et al, 1987; Saccone et al, 1990).

This method of Jukes and Cantor was developed under the assumption that the rate of nucleotide substitution was the same for all pairs of the four nucleotides A, T, G and C (i.e. a random process), and gives the maximum likelihood estimate of the number of nucleotide substitutions between two sequences (Jukes et al, 1969). Jukes-Cantor is a one parameter distance method and does not take into account differences in the frequency of transitions (purine to purine and pyrimidine to pyrimidine substitutions) and transversions (purine to pyrimidine substitutions) when calculating evolutionary distances. However, in this study, the pairwise distances between sequences were relatively small and it is unlikely that saturation will have occurred at sites of transition to an extent that major underestimates of the rate of sequence change would have been made. This one parameter distance method can be used to calculate distances for synonymous (do not result in amino acid changes) and non-synonymous (change amino acids) substitutions as there is no transition/transversion bias in this analysis.

The assumption that all nucleotide substitutions occur randomly, as in the Jukes-Cantor model, is unrealistic in some cases as transitions generally occur more frequently than transversions. Kimura proposed a two parameter model in which the rate of transitional and transversional substitutions per site are taken into account (Kimura, 1980).
1.6.2 RATES OF NUCLEOTIDE SUBSTITUTION.

The rate of nucleotide substitution is defined as the number of substitutions per site per year and can be calculated by dividing the pairwise distance calculated for any two sequences with twice the time of divergence between the two sequences. The divergence time is assumed to be the same for the two sequences considered. The rate of nonsynonymous substitution may be extremely variable among genes as the majority of nonsynonymous substitutions are subject to purifying selection from external factors such as the host immune response. Consequently, nonsynonymous substitutions which improve protein function may be selected resulting in a greater rate of nonsynonymous to synonymous substitution. Reciprocally, nonsynonymous substitutions resulting in a deleterious effect on protein function will be eliminated by selection, reducing the rate of nonsynonymous substitution. However, synonymous substitutions do not cause changes in amino acids and will not be subject to purifying selection. These substitutions are thought to reflect the underlying mutational rate allowing the estimation of the rate of nucleotide substitution.
1.6.3 ESTIMATION OF DIVERGENCE TIMES.

Sequence data can be used to estimate times of divergence between species, individuals and indeed sequences within an individual. Assuming the rate of nucleotide substitution (evolution) is known from a previous study (see section 1.6.2) the time of divergence between two sequences can be estimated by dividing the synonymous pairwise distance by twice the rate of change. Again this calculation assumes rate constancy, an assumption which often does not always hold. Using synonymous pairwise distances reduces any errors incurred while estimating divergence times as these nucleotide sites are not subject to selection pressures to the extent that nonsynonymous substitutions are. Li et al. have proposed a method that can reduce the effects of unequal rates of substitution when estimating times of divergence (Li et al, 1987).

1.6.4 MOLECULAR PHYLOGENY.

In phylogenetic studies, the evolutionary relationship among a group of organisms (population tree) or genes (gene tree) can be illustrated by means of a phylogenetic tree. Phylogenetic trees can be either rooted or unrooted. A rooted tree indicates the direction of evolution, and the root is the common ancestor of all the OTUs (operational taxonomic units) studied. An unrooted tree specifies the relationship among the OTUs but does not define the evolutionary path. In practice
the majority of tree-making methods produce unrooted trees. However, in order to root an unrooted tree an outgroup can be added which is evolutionarily related to the OTUs under study, having diverged from the other OTUs prior to their divergence from one another.

There are a number of methods available for the construction of phylogenetic trees (Nei, 1987; Felsenstein, 1988), which can be classified as distance matrix methods and character based (maximum parsimony) methods. Distance matrix methods involve the calculation of evolutionary distances (pairwise distances) for all pairs of taxa (e.g. nucleotide sequences), and a phylogenetic tree is constructed by certain principles and algorithms. In maximum parsimony methods discrete character states (e.g. nucleotides or amino acids at a site) are used, and a phylogenetic tree is constructed by considering the shortest pathway leading to these character states.

The simplest method for reconstruction of phylogenetic trees is the unweighted pair group method with arithmetic mean (UPGMA). This method employs a sequential clustering algorithm. Initially the two OTUs that are most similar from among all the OTUs are identified and subsequently are treated as a new single OTU. This new OTU is referred to as a composite OTU. Subsequently, the next pair with the highest similarity are identified, and so on, until only two OTUs are left. This method assumes a constant rate of evolution and hence a rooted tree is produced.

Computer simulations have shown that one of the most efficient distance methods for producing phylogenetic trees with the correct topology (branching
pattern of a tree) is the neighbor-joining method (Nei et al, 1991) proposed by Saitou and Nei (1987). This method is a simplified version of the minimum evolution method in which distance measures that correct for multiple hits at the same sites are used. In this method the smallest value of the sum of all branches is chosen as the correct tree. In the neighbor-joining method the smallest value of the sum of all branches is not calculated for all the topologies. Instead, the examination of different topologies is embedded in the algorithm, so that only one final tree is produced. This method produces an unrooted tree, and usually requires an outgroup to find the root.

The principle of maximum parsimony or minimum evolution involves the identification of a tree that requires the smallest number of evolutionary changes to explain the differences observed among the OTUs being studied. However, quite often more than one tree with the same minimum number of changes can be found, so that no unique tree may be inferred in some instances. These topologies are called equally parsimonious trees. When constructing a maximum parsimony tree informative sites must first be identified. These are sites at which at least two different kinds of nucleotides, each represented twice, are used. Other variable sites are not used for constructing maximum parsimony trees, although they are informative for distance methods. Following this the minimum number of substitutions at each informative site is calculated for each possible tree. Finally, the sum of the number of changes over all the informative sites is calculated for each tree and the tree associated with the smallest number of substitutions is chosen. The maximum parsimony method produces unrooted trees, primarily
determining the topology of a tree.

1.6.5 STATISTICAL SIGNIFICANCE OF TREE TOPOLOGIES.

There are two different types of methods available for testing the reliability of a tree obtained. These are the maximum likelihood method (Felsenstein, 1981) and the bootstrap test (Felsenstein, 1985). The maximum likelihood method examines the reliability of every interior branch of the tree. This method can be very time consuming and is generally limited to small data sets. The bootstrap test involves randomly resampling the data from which the tree was constructed, producing a new tree with the resampled data. This process is repeated several hundred times and the frequency at which particular branches are observed in the newly constructed trees is calculated giving a probability to each branch in the original tree. This statistical method is suitable for use in neighbor-joining and maximum parsimony methods.
AIMS.
The aims of this study were to characterize genetically HIV-1 variants obtained from various lymphoid and non-lymphoid tissues, investigate whether separate populations existed and determine whether such differences reflected variation in biological properties, such as their inferred in vitro tropism and cytopathology. I also planned to investigate the diversity of variants (LPA) in pre-symptomatic and symptomatic patients to identify whether more or less divergent populations emerge during disease progression. Finally, I have investigated whether, through evolutionary analysis of the p17_{gag} region, the time of spread of HIV-1 from lymphoid to non-lymphoid tissues can be estimated to determine if this occurs as an early or late event in disease progression.
CHAPTER TWO: MATERIALS AND METHODS.
2.1 CLINICAL DETAILS OF STUDY PATIENTS.

Tissues from various organs were obtained at autopsy, carried out within three days of death, from 43 HIV-infected individuals. Eight individuals died from other reasons and were classified as pre-symptomatic on the basis of an absence of HIV-related symptoms before death and an absence of HIV-related changes in pathological examination. In contrast, 35 individuals died from complications associated with HIV-infection, such as opportunistic infections and neoplasia (symptomatic group). Details of the clinical symptoms each patient presented with are summarised in Table 5. Samples of brain (left frontal lobe), lymph node (mesenteric) from each individual were dissected into 1 to 2 cm pieces and stored at -70°C. Samples of lung, colon and spleen were also analysed from patient 4. CD4 lymphocyte counts were obtained on three occasions in the year before death from each individual (Table 5).
### TABLE 5.

Abbreviations: IVDU: injecting drug abuser; HO: male homosexual; HE: heterosexual contact; B: infection through blood transfusion

CD4 lymphocyte (cell/μl) count over one year prior to death (mean of 2-3 values).
N/D=not done.

Diagnosis of giant cell encephalitis, based upon pathology, p24 antigen detection and quantitative PCR (Bell, 1996).

Abbreviations: 1=Pneumocystis carinii pneumonia; 2=oral thrush; 3=oesophageal thrush; 4=oral candidiasis; 5=oesophageal candidiasis; 6=dementia; 7=shingles; 8=neurological degeneration; 9=atypical mycobacterial infection; 10=herpes zoster; 11=oral ulceration; 12=lymphoma; 13=cerebral atrophy; 14=widespread cytomegalovirus infection; 15=lymphadenopathy; 16= splenomegaly; 17=cutaneous Kaposi’s sarcoma; 18=systemic Kaposi’s sarcoma; 19=dysphasia; 20=cerebral toxoplasmosis; 21=autonomic neuropathy; 22=persistent diarrhoea; 23=weight loss; 24=muscle wasting; 25=HIV retinopathy; 26=HIV myelitis; 27=focal cytomegalovirus infection in brain.

This patient is the heterosexual contact of a known IVDU.

These patients are included in the study described in chapter 3.

This patient is not included in the study described in chapter 4.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Age</th>
<th>Risk</th>
<th>CD4 count</th>
<th>GCE</th>
<th>Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>76</td>
<td>41</td>
<td>IVDU</td>
<td>802</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>31</td>
<td>IVDU</td>
<td>370</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>32</td>
<td>IVDU</td>
<td>240</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>31</td>
<td>IVDU</td>
<td>180</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>28</td>
<td>IVDU</td>
<td>170</td>
<td>-</td>
<td>20</td>
</tr>
<tr>
<td>68</td>
<td>28</td>
<td>HE</td>
<td>155</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>23</td>
<td>HE</td>
<td>140</td>
<td>-</td>
<td>1,2</td>
</tr>
<tr>
<td>2</td>
<td>46</td>
<td>IVDU</td>
<td>140</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>27</td>
<td>HE</td>
<td>115</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>28</td>
<td>HO</td>
<td>105</td>
<td>+</td>
<td>1,21</td>
</tr>
<tr>
<td>5f</td>
<td>34</td>
<td>IVDU</td>
<td>95</td>
<td>+</td>
<td>2,6</td>
</tr>
<tr>
<td>69</td>
<td>28</td>
<td>IVDU</td>
<td>90</td>
<td>+</td>
<td>4,13,23,25,26</td>
</tr>
<tr>
<td>25</td>
<td>38</td>
<td>IVDU</td>
<td>85</td>
<td>-</td>
<td>12,19</td>
</tr>
<tr>
<td>40</td>
<td>33</td>
<td>IVDU</td>
<td>50</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>42</td>
<td>IVDU</td>
<td>40</td>
<td>-</td>
<td>15</td>
</tr>
<tr>
<td>19</td>
<td>35</td>
<td>IVDU</td>
<td>33</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>25</td>
<td>IVDU</td>
<td>28</td>
<td>+</td>
<td>6,11</td>
</tr>
<tr>
<td>79f</td>
<td>28</td>
<td>IVDU</td>
<td>26</td>
<td>+</td>
<td>11,12,13,14</td>
</tr>
<tr>
<td>24</td>
<td>36</td>
<td>HO</td>
<td>19</td>
<td>-</td>
<td>1,9,13,16,17</td>
</tr>
<tr>
<td>44</td>
<td>32</td>
<td>HO</td>
<td>19</td>
<td>-</td>
<td>12,13</td>
</tr>
<tr>
<td>36</td>
<td>50</td>
<td>HO</td>
<td>15</td>
<td>-</td>
<td>1,14,21</td>
</tr>
<tr>
<td>54</td>
<td>25</td>
<td>IVDU</td>
<td>13</td>
<td>+</td>
<td>27</td>
</tr>
<tr>
<td>17</td>
<td>24</td>
<td>IVDU</td>
<td>12</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>32</td>
<td>30</td>
<td>IVDU</td>
<td>11</td>
<td>+</td>
<td>15,24</td>
</tr>
<tr>
<td>18</td>
<td>44</td>
<td>HO</td>
<td>10</td>
<td>-</td>
<td>1,2,3,9,17</td>
</tr>
<tr>
<td>51</td>
<td>35</td>
<td>HO</td>
<td>10</td>
<td>+</td>
<td>12,15</td>
</tr>
<tr>
<td>4f</td>
<td>34</td>
<td>IVDU</td>
<td>8</td>
<td>+</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>34</td>
<td>IVDU</td>
<td>8</td>
<td>+</td>
<td>1,5,13</td>
</tr>
<tr>
<td>10</td>
<td>33</td>
<td>IVDU</td>
<td>6</td>
<td>-</td>
<td>13,15,16</td>
</tr>
<tr>
<td>21</td>
<td>30</td>
<td>B</td>
<td>6</td>
<td>+</td>
<td>12,13,20</td>
</tr>
<tr>
<td>15</td>
<td>34</td>
<td>IVDU</td>
<td>5</td>
<td>+</td>
<td>1,4,5,9</td>
</tr>
<tr>
<td>27</td>
<td>31</td>
<td>HO</td>
<td>5</td>
<td>-</td>
<td>13,14</td>
</tr>
<tr>
<td>45</td>
<td>45</td>
<td>HO</td>
<td>5</td>
<td>-</td>
<td>1,9,20</td>
</tr>
<tr>
<td>16</td>
<td>30</td>
<td>HO</td>
<td>4</td>
<td>-</td>
<td>2,12,14,17,18</td>
</tr>
<tr>
<td>77</td>
<td>34</td>
<td>HO</td>
<td>3</td>
<td>-</td>
<td>1,13,14</td>
</tr>
<tr>
<td>63</td>
<td>46</td>
<td>IVDU</td>
<td>2</td>
<td>+</td>
<td>13,14</td>
</tr>
<tr>
<td>37</td>
<td>30</td>
<td>HO</td>
<td>2</td>
<td>-</td>
<td>9,22</td>
</tr>
<tr>
<td>62</td>
<td>36</td>
<td>IVDU</td>
<td>2</td>
<td>-</td>
<td>1,4,9,23</td>
</tr>
<tr>
<td>64</td>
<td>36</td>
<td>HO</td>
<td>2</td>
<td>-</td>
<td>4,9,10,23</td>
</tr>
<tr>
<td>30</td>
<td>26</td>
<td>IVDU</td>
<td>0</td>
<td>+</td>
<td>1,10</td>
</tr>
<tr>
<td>78</td>
<td>33</td>
<td>IVDU</td>
<td>0</td>
<td>-</td>
<td>1,14</td>
</tr>
<tr>
<td>6f</td>
<td>28</td>
<td>HE</td>
<td>N/D</td>
<td>+</td>
<td>7,12</td>
</tr>
<tr>
<td>38</td>
<td>43</td>
<td>HO</td>
<td>N/D</td>
<td>-</td>
<td>9,14,24</td>
</tr>
</tbody>
</table>
2.2 BUFFERS AND REAGENTS.

2.2.1 EXTRACTION OF DNA FROM TISSUE SAMPLES.

Lysis buffer: 50nM Tris hydrochloride (pH 8.0)
   100mM NaCl
   50mM EDTA
   1% sodium-n-lauroylsarcosine
   100µg of proteinase K per ml

Phenol (Rathburn chemicals Ltd)

Chloroform (AnalaR)

100% Ethanol (AnalaR)

Distilled water

Centrifuge (Hereus Biofuge 15R)

2.2.2 AMPLIFICATION OF PROVIRAL DNA.

10 X PCR buffer: 200mM Tris-HCL (pH 8.8)
   500mM KCL
   15mM MgCl₂
   0.5% Triton X-100
Nucleoside triphosphate mixture (100mM stock; dilute to 3mM; Boehringer Mannheim)

*Taq* polymerase (Promega: $1 \mu l$ is equal to 5 units)

Water cooled thermal cycler (Techne PHC-1A and GeneE)

2% agarose gel: 6 g agarose (Sigma)

300 ml 10 x TBE

20 $\mu l$ Ethidium Bromide

10 x TBE (1 litre): 108 g Tris base (AnalaR)

55 g Boric acid (Molecular Biology Certified; Kodak)

40 ml 0.5M EDTA (Molecular Biology Certified; Kodak)

Make up to 1L with dH$_2$O

1 x TBE buffer used as electrophoresis buffer in gel tank.

**2.2.3 SOLID PHASE PURIFICATION OF PCR PRODUCTS.**

BSA: 0.1% bovine serum albumin in PBS (pH 7.2).

PBS (pH 7.5): 137 mM NaCl

2.7 mM KCL

4.3 mM Na$_2$HPO$_4$.7H$_2$O

1.4 mM KH$_2$PO$_4$
Binding and washing buffer (BW): 10 mM Tris-HCL (pH 7.5)

1 mM EDTA

2.0 M NaCl (final concentration 1.0 M)

TE buffer: 10 mM Tris-HCL (pH 7.5)

1 mM EDTA

0.15 M NaOH

Dynabeads M-280 streptavidin (Dynal).

Dynal magnetic particle concentrator 6 (MPC 6; Dynal).

2.2.4 SEQUENCE ANALYSIS.

5 x Sequenase reaction buffer: 200 mM Tris-HCL (pH 7.5)

100 mM MgCl₂

250 mM NaCl

5 x labelling mix: 7.5 μM dGTP

7.5 μM dTTP

7.5 μM dCTP

[α³⁵S] dATP or [α-³²P] dATP (Amersham).

Termination solutions: 80 μM of all four dNTPs supplemented with 8 μM ddATP,
ddTTP, ddGTP or ddCTP.

Stop solution: 95% formamide
20 mM EDTA
0.05% xylene cyanol
0.05% bromophenol blue

5% Denaturing PAGE gel: 21 g Urea (AnalaR)
5 ml Long Ranger gel solution (50%; Flowgen)
5 ml 10 X TBE (Sanger)
0.05 g Ammonium persulfate (APS; Sigma)
Make up to 50 ml with dH₂O
Add 25 μl TEMED prior to pouring gel mix.

10 x TBE (Sanger; 2 litre): 324 g Tris base
85 g Boric acid
19 g EDTA
Make up to 2L with dH₂O.
2.3 EXTRACTION OF DNA FROM TISSUE SAMPLES.

Extraction of DNA from tissues was carried out by re-suspending 1-2 cm pieces of tissue in 500 µl of lysis buffer and incubated at 65°C for two hours. Following this 450 µl of phenol was added to the extraction tubes and mixed thoroughly. Extraction tubes were then centrifuged at 13000 rpm for 10 mins. This process was repeated until the interface was clear. The aqueous layer was then transferred to a fresh 1.5 ml eppendorf containing 450 µl of chloroform. This mixture was then mixed thoroughly and centrifuged at 13000 rpm for 10 mins. The aqueous layer was again transferred to a fresh 1.5 ml eppendorf containing 900 µl of 100% ethanol, mixed and left overnight at -20°C to allow precipitation of the nucleic acid. The nucleic acid extract was collected by centrifugation at 13000 rpm at 0°C for 20 min. The supernatant was then discarded and the precipitate dried at 65°C for 10 min. Following this the nucleic acid pellet was re-suspended in 100-200 µl of distilled water. To ensure adequate solubilization of the pellet the sample was incubated at 50°C for 15-30 min. During each extraction process 1-2 cm pieces of known positive and negative tissue samples were included providing a comprehensive check for contamination.

2.4 QUANTITATION OF TOTAL EXTRACTED DNA.

The concentration of DNA in each sample was accurately quantified using spectrophotometric UV absorbance readings at wavelengths of 260 and 280 nm.
Samples were diluted 1 in 20 in pyrogen free water and the absorbance measured in a spectrophotometer. The ratio of the optical density (OD) at 260 nm to that at 280 nm gave an indication of purity of each preparation. Pure DNA preparations have an $A_{260}/A_{280}$ ratio of approximately 1.8. The concentration of DNA was calculated from the following equation: $A_{260} \times D \times 50 = \text{DNA concentration}$ ($\mu g/ml$), where $A_{260} =$ absorbance at 260nm, $D$ is the dilution factor (=20) and 50 is equivalent to the concentration ($\mu g/ml$) of ds DNA at $A_{260}$ of 1.0.

2.5 AMPLIFICATION OF PROVIRAL DNA.

Proviral DNA was amplified using the polymerase chain reaction (PCR). Nested PCR was employed, further amplifying the first PCR product with a second set of specific primers that lie within those employed in the first PCR. This increases the overall sensitivity and specificity of the PCR, increasing the amplification by 10 000 fold compared to that attained by using only one set of primers. The first and second PCR were subjected to a thermal cycle of 36 seconds at 94°C to allow denaturation of ds DNA, 42 seconds at 46°C (V1/V2) and 50°C (gag), to allow primer annealing to the ss DNA and 90 seconds at 72°C to allow strand extension. Each template strand was subjected to 25 cycles of amplification. At the end of the last cycle samples were heated to 72°C for 5 min to allow termination of uncompleted stands. The first PCR was carried out in a 50 $\mu$l volume, for each sample, containing 5 $\mu$l of DNA, 5 $\mu$l of 10x PCR buffer, 0.5 $\mu$l of nucleoside triphosphate mixture (3 mM each of dGTP, dATP, dTTP and dCTP),
0.25 μl of sense primer (approximately 20 μM), 0.25 μl of antisense primer (approximately 20 μM), 38 μl of pyrogen free water and 1 unit of Taq polymerase. The second, nested, PCR was carried out in a 20 μl volume, for each sample, containing 1 μl of primary PCR product, 2 μl of 10 x PCR buffer, 0.2 μl of dNTP's, 0.1 μl of sense primer (approximately x 8 μM), 0.1 μl of antisense primer (approximately x 8 μM), 16.5 μl of pyrogen free water and 0.4 units of Taq polymerase. Prior to transfer to the thermal cycler each sample was covered with a drop of mineral oil to prevent loss of sample due to evaporation. Amplification of DNA was carried out using primers flanking hypervariable regions 1 and 2 from env and p17 from gag. The nucleotide sequences of the primers were as follows:

V1/V2:

a: GAG GAT ATA ATC AGT TTA TGG; + (sense), 6539  
b: GA TCA AAG CCT AAA GCC ATG; +, 6560  
c: TTG AAA GAG CAG TTT TTT ATC TCT CC; - (antisense), 6677  
d: TG(A)A AAA ACT GCT CTT TCA A; +, 6684  
e: CAA TAA TGT ATG GGA ATT GG; -, 6857  
f: AAT GTA CTG TGC TGA CAT T; -, 6944

GAG:

a: GCG AGA GCG TCA GTA TTA AGC GG; +, 795  
b: GGG AAA AAA TTC GGT TAA GGC C; +, 835  
c: CTT CTA CTA TTT TTA CCC ATG C; -, 1248  
d: TCT GAT AAT GCT GAA AAC ATG GG; -, 1296
All positions numbered according to the HXB2 genome: (Myers et al., 1991)

During each PCR amplification 1 µg of known positive and negative DNA samples were included providing a comprehensive check for contamination. An additional negative control was included whereby no DNA was added to the reaction mix ensuring contamination of the PCR buffer had not occurred.

2.6 LIMITING DILUTION ASSAY.

In order to obtain single molecules of HIV proviral DNA to allow analysis of heterogenous viral sequences a limiting dilution assay was carried out. To isolate single proviral sequences multiple replicates of DNA were amplified at a dilution resulting in the detection of only a small proportion of positive PCR products. Using the poisson formula, which assumes a random distribution of DNA molecules, the probability of positive PCR products having originated from a single proviral sequence can be calculated. If 20% of the PCR replicates are positive then approximately 95% of reactions can be said to have originated from a single proviral sequence. Initially triplicates of 4 ten-fold dilutions were set up (0.1 µg-0.0001 µg) and amplified using nested PCR, as previously described. A further 20 replicates were set up at the dilution resulting in the least number of positive reactions and amplified using nested PCR. If 20% of these replicates proved to be positive they were considered to have originated from a single proviral sequence and were retained for sequence analysis.
2.7 VISUALIZATION OF AMPLIFIED PCR PRODUCTS.

Amplified PCR products were visualized on 2% agarose gels containing ethidium bromide. Ethidium bromide is an intercalating agent which exhibits a fluorescence under UV light, allowing any positive PCR products to be detected as fluorescent bands. The gel was run for 10 mins at 150 V and analyzed under UV light.

2.8 SOLID PHASE PURIFICATION OF PCR PRODUCTS.

For direct sequencing of PCR products a higher purity than that provided by previous ethanol precipitation is required. The sequencing reaction (Sanger method) is highly sensitive to various components present in the PCR product, such as dNTP's, primers and buffer solutions. Single strand purification was therefore carried out using a solid phase method (Dynal) allowing highly specific isolation and purification of target PCR products. This purification method involves a highly specific interaction between streptavidin (covalently attached to the bead surface) and biotin (incorporated into the primer oligonucleotide). Therefore, prior to purification, a secondary PCR reaction was performed using one biotinylated primer and one unlabelled primer generating a PCR product with one strand having a biotin moiety at either the 5' or 3' end. This reaction is carried out in a 100 μl volume, for each sample, containing 1 μl of PCR product, 10 μl of reaction buffer, 1 μl of dNTPs (approximately 33 mM ), 0.25 μl of normal primer (approximately
20 mM), 0.25 μl of biotinylated primer (approximately 20mM), 87.1 μl of pyrogen free water and 0.4 μl of Taq (2 units per reaction). For each sample to be sequenced 20 μl of re-suspended Dynabeads were transferred to a 1.5 ml eppendorf and washed with 40 μl of 0.1% bovine serum albumin. This was then placed in a magnet (Dynal MPC-6) and the supernatant removed. The beads were then re-suspended in 40 μl of BW, mixed with 40 μl of the amplified PCR product and incubated at room temperature for 20 mins, re-suspending the beads occasionally, allowing the immobilization of the PCR product to the Dynabeads. The beads were then washed in 40 μl of BW and re-suspended in 8 μl of 0.15 M NaOH and incubated at room temperature for 10 mins, denaturing the ds DNA bound to the beads leaving only ss DNA bound. The supernatant, containing the unbound complementary sense DNA, was then removed and the beads were washed once with 50 μl of NaOH, once with 50 μl of BW and once with 50 μl of TE, ensuring the complete separation of the DNA strands. Finally, the beads were re-suspended in 20 μl of TE and stored at 4°C. During washes beads must only be pipetted up and down gently and should not be centrifuged or vortexed.

2.9 DIRECT SEQUENCING OF AMPLIFIED PCR PRODUCTS.

Single molecules of HIV provirus were isolated by limiting dilution and amplified in a nested PCR to produce sufficient DNA to allow direct sequencing of the PCR products. Direct sequencing of amplified DNA was achieved by using a solid phase sequencing method described above. In this study Sequenase Version
2.0 DNA sequencing kits were used (United States Biochemical) with Sequenase™ Version 2.0 T7 DNA polymerase. This enzyme is a genetic variant of bacteriophage T7 DNA polymerase created by \textit{in vitro} genetic manipulation, completely lacking a 3' to 5' exonuclease activity and including properties such as high processivity, high speed and the ability to use nucleotide analogues for sequencing making it the ideal enzyme to use for chain termination sequencing. For each template strand of ssDNA a single annealing reaction was used to allow the primer strand to bind to the ssDNA. The primer is needed because template dependent DNA polymerases are unable to initiate DNA synthesis on an entirely single stranded molecule. A short double stranded region is required to provide a 3' end onto which the polymerase can add new nucleotides, annealing of the primer provides this 3' end without which strand synthesis would not take place. For each template, 2 \( \mu l \) of the appropriate primer, 2 \( \mu l \) of 5 x Sequenase reaction buffer, 1 \( \mu l \) dimethylsulphoxide (DMSO) and 5 \( \mu l \) of DNA (bound to magnetic beads) were mixed together. This mixture was then warmed at 65°C for approximately 5 mins and allowed to cool slowly to room temperature over a period of approximately 30 mins. To each annealed template-primer the following labelling (extension) mix was added: 1 \( \mu l \) dithiothreitol (DTT, 0.1 M), 2 \( \mu l \) of diluted labelling nucleotide mix (1:20), 0.5 \( \mu l \) of \([\alpha-^{35}S]\) or \([\alpha-^{32}P]\) dATP (5 \( \mu Ci \)) and 2 \( \mu l \) diluted Sequenase polymerase (1:8, approximately 3.25 units). The annealing reaction and labelling mix were mixed thoroughly and incubated at room temperature for approximately 5 mins, during which time the termination reaction was prepared. For each template four tubes were labelled 'G', 'A', 'T', and 'C'. Each tube was filled with 2.5 \( \mu l \) of
the appropriate dideoxy termination mixture and pre-warmed at 37°C (the optimal
temperature for Sequenase polymerase enzyme). When the labelling reaction was
complete, 3.5 μl were transferred to each of the four termination tubes (G, A, T
and C) and incubated at 37°C for approximately 5 mins and then 4 μl of stop
solution was added to each termination reaction stopping the termination reaction.
Following this sequencing reaction the product was heated to 95°C to destroy the
interaction between the streptavidin and biotin and release the DNA templates
together with the dideoxy fragments. This product was then electrophoresed on a
5% denaturing polyacrylamide gel (PAGE). Gels were run for 2 hours allowing
approximately 250 to 300 bases to be read. Gels were dried and exposed overnight
on BioMAX film.

2.10 LENGTH PROFILING ANALYSIS OF AMPLIFIED PCR PRODUCTS.

To obtain length profiles across the V1 and V2 hypervariable regions
proviral DNA was amplified using nested PCR, as previously described. However,
the second PCR was modified as detailed below. The concentration of the dNTP’s
were reduced from 33 μM to 8 μM with the addition of 2 μCi[α35-S]
dATP/reaction. In addition 0.1 μl instead of 1 μl of primary PCR product was
transferred reducing extra bands resulting from the carry over of excess primary
PCR product and primers. Subsequently, 1μl of the secondary PCR product was
mixed with 5 μl of stop solution and heated to 90°C to denature DNA. 5-6 μl were
then electrophoresed on a 5% denaturing polyacrylamide gel. Gels were run for
approximately 30 mins. Gels were then dried and exposed overnight on BioMAX film.

At the start of this study [$\alpha^{35}$-S]dATP was used. However, it has recently been shown that this radionucleotide releases $H_2S$ when heated during thermal cycling which can permeate polypropylene tubes and cause radioactive contamination. In more recent experiments this radionucleotide has been substituted with [$\alpha^{32}$-P]dATP in all experiments, using 0.5 $\mu$Ci per reaction.

2.11 NUCLEOTIDE SEQUENCE ACCESSION NUMBERS.

The sequences obtained in this study have been submitted to GenBank and assigned accession numbers U79785 to U79869 (gag; P4, 5 and 6) and U79870 to U80057 (V1/V2; P4, 5 and 6). Nucleotide sequences for P79 in the p17 $\text{gag}$ region are listed in appendix IV.
CHAPTER 3: INVESTIGATION OF THE DYNAMICS OF THE SPREAD OF HUMAN IMMUNODEFICIENCY VIRUS TO BRAIN AND OTHER TISSUES BY EVOLUTIONARY ANALYSIS OF SEQUENCES FROM THE p17 gag AND env GENES.
3.1 INTRODUCTION.

Infection with HIV is associated with a slow, progressive and irreversible impairment of the immune system eventually leading to AIDS. Inherent in the nature of infection with HIV-1 is the prolonged asymptomatic period that proceeds the development of disease (Fauci, 1993b; Miedema et al., 1990b; Bednarik et al., 1992b; Pantaleo et al., 1993b), where infection may be subclinical for as long as 10-15 years. This phenomenon was originally hypothesised to result from viral latency, whereby viral or proviral DNA became integrated into the host genome with the simultaneous cessation of viral expression and independent replication (Bednarik et al., 1992). The ensuing progression to AIDS would then result from subsequent re-activation of virus replication by various factors acting upon the infected cell such as antigens, mitogens and transcriptional factors produced by other viruses. However, it has been recently shown that from the time of seroconversion there is active replication of the virus in lymphoid tissues (Fauci, 1993b; Miedema et al., 1990b; Pantaleo et al., 1993b). There are few convincing demonstrations of active infection of non-lymphoid tissues until later in infection, and this change in distribution may be associated with increased immunosuppression in AIDS (Donaldson et al., 1994b). Alternatively, it is possible that variants detected in non-lymphoid tissue such as brain in AIDS patients may have been continuously present from initial infection, but that infection only becomes clinically significant during severe immunosuppression. In this model, HIV encephalitis could be regarded as re-activation rather than de novo infection.
This study was undertaken to estimate the time of spread of HIV-1 to non-lymphoid tissues to determine whether re-activation or actual virus spread was responsible for the pathology observed in non-lymphoid tissues in AIDS. The individuals observed in this study were from a cohort of infected IDU identified in Edinburgh. These individuals were chosen for this study because infection of this cohort occurred at approximately the same time (Robertson et al., 1986a), and clinical follow up studies have been possible due to the isolation of this group of individuals and their social stability (Goodwin et al., 1996; Davies et al., 1995; Brettle et al., 1996). Serologic analysis of stored serum samples indicated that HIV-1 was introduced into Edinburgh and spread rapidly in late 1982 and early 1983 (Peutherer et al., 1985). This rapid spread of the virus in Edinburgh may be explained by high risk behavioural habits of IDU in Edinburgh. At the time of this rapid spread there was a high frequency of needle sharing and use of contaminated needles (Robertson et al., 1986b; Robertson et al., 1986b), the likely result of difficulties in obtaining sterile needles and syringes. A recent study by Holmes et al., analysed the phylogenetic relationships between HIV-1 infected individuals in different risk groups in Edinburgh (Holmes et al., 1995). Phylogenetic analysis using the p17 region from the gag gene revealed three distinct groups within this population, two of which comprised sequences from haemophiliac patients and a third distinct group containing sequences from IDU and those infected from them through heterosexual contact. This finding was somewhat surprising, suggesting that the epidemic in this group may have arisen from a single HIV variant that first infected the IDU population in 1982/83. Therefore, this IDU cohort provides a
unique opportunity to examine the divergence of viral variants over time within a population infected from a common source.

In this study sequences were obtained from the p17$_{ gag }$ region and V1/V2 flanking regions of HIV-1 from four individuals from the Edinburgh IDU cohort. The p17$_{ gag }$ region was chosen because most nucleotide differences in this region are synonymous and therefore are not subject to positive selection pressures, unlike many other regions in the HIV-1 genome. Nonsynonymous substitutions alter the amino acid composition and may arise through a process of natural selection. For example, the appearance of escape mutants through selection pressures exerted by the host immune system. However, synonymous substitutions do not cause amino acid changes of proteins and are therefore less subject to selection. Hence, synonymous substitutions are more likely to be caused by a random fixation of mutations (neutral theory) and are therefore more likely to reflect the underlying mutational rate allowing reliable estimates of divergence times to be calculated.

It has previously been suggested that the extent of HIV-1 replication throughout infection is not constant. Extensive replication was previously thought to occur during the acute phase of infection and again later on in infection prior to the progression to AIDS. These observations have led to the suggestion that the rate of sequence change may be greatly increased early and late in infection and greatly reduced during the asymptomatic phase. However, more recent studies have shown continuous productive viral replication, regardless of clinical stage, with $10^8$ to $10^9$ virions produced daily (Connor et al., 1994; Ho et al., 1995; Wei et al., 1995). Secondly, more extensive replication does not necessarily imply that the rate
of sequence change will occur more rapidly, being principally dependent on the replicative cycle of the virus, rather than immune control of the host. If rates of sequence change are measured at genuinely neutral sites, then the observed rate should simply be proportional to the number of replicative cycles.

The rate of sequence change of the p17_gag region has previously been determined from haemophiliacs, in the first two years of HIV-1 infection, who were infected from a common batch of clotting factor concentrate (Kasper et al., 1995). Greater rates of synonymous (6.0 to 7.2 x 10^{-3}) to nonsynonymous substitution (3.6 to 4.8 x 10^{-3}) were observed, suggesting that this region is not subject to positive selection pressures. In this study, of individuals infected for 10 years or more, similar rates of synonymous and nonsynonymous substitutions were observed, suggesting that the rate of sequence change for the p17 region of gag is constant from early on in infection (2 years) until AIDS (10 years). This region is therefore of use in reconstructing epidemiological relationships between HIV-1 infected individuals (Holmes et al., 1993) and can be used as a type of molecular clock allowing estimates of the divergence time between any two sequences to be carried out. This molecular clock can be extended to the comparison of variants within different cell types within a single infected individual.

Tissues from various lymphoid and non-lymphoid organs were obtained at autopsy from a number of HIV-1 positive patients known to have a high viral load in the brain and evidence of giant cell encephalitis (GCE) by pathology. Phylogenetic analysis of both p17_gag and V1/V2 flanking regions was carried out in order to explore the relationship between the various lineages present and the
spread of infection to non-lymphoid tissues. It was possible to estimate the time of
divergence between lymphoid and non-lymphoid tissues allowing an estimation of
the length of time, prior to death, that non-lymphoid tissue has been infected.

3.2 CLINICAL DETAILS OF PATIENT SAMPLES.

Tissue from various organs were obtained at autopsy, carried out within
three days of death, from four individuals who died with an AIDS-defining illness.
All individuals showed evidence of HIV infection of the brain upon post-mortem
examination, as determined by the histological appearance of giant cells, the
detection of p24 by immunocytochemistry, and the finding of high proviral loads in
brain by quantitative PCR (Donaldson et al., 1994b). Pathological examination of
the fixed brains revealed evidence of atrophy on external examination and this was
confirmed on section by the presence of ventricular dilatation and opening up of
the sulci in all four patients. In p6 a focal 1cm diameter lesion was identified on
macroscopic inspection in the right basal ganglia. Histological examination of this
lesion showed that it was a primary CNS lymphoma. In p79 focal 2cm diameter
lesions were identified upon macroscopic inspection in the left thalamus, internal
capsule and left occipital lobe both in cortical ribbon and paraventricular regions.
Microscopic examination showed that these lesions were composed of
periventricular foci of lymphoma, much of which was necrotising. Neither of the
other two patients showed macroscopic focal lesions of the brain. Histological
examination in all four patients displayed evidence of quite florid HIV encephalitis
and leukoencephalopathy, characterised by giant cells and focal collections of macrophages and microglia cells, associated with myelin damage. There was no evidence of perivascular or leptomeningeal inflammatory infiltrates and in particular, lymphocytes were not identified within the CNS parenchyma. Additional clinical information for the five years prior to death for each patient is summarised in Fig 6.

3.3 PHYLOGENETIC ANALYSIS.

Sequence comparisons between viruses from the four study patients were made in the p17\textsubscript{gag} gene and in three of the study patients (p4, 5 and 6), in the hypervariable flanking regions of V1 and V2 of the env gene of HIV-1. The p17\textsubscript{gag} region amplified began at nucleotide 795 of HXB2 and extended to position 1319. The V1/V2 region amplified began at position 6539 of HXB2 and extended to position 6976. The length of the gag region used for sequence comparisons was 413 nucleotides and that of the V1/V2 region was 297 nucleotides. All nucleotide sequences were aligned manually. An unrooted phylogenetic tree for 85 p17\textsubscript{gag} (p4, 5 and 6) nucleotide sequences obtained from lymph node, brain and lung samples was constructed using the neighbour-joining method using the program NEIGHBOR in the PHYLIP package (version 3.5; Felsenstein, 1989). Distances between each pair of sequences were estimated using the program DNADIST in the PHYLIP package (version 3.5; Felsenstein, 1989). Rooted trees were constructed for each patient using the MEGA package with the sequence of HIV\textsubscript{MN}
Fig. 6. Clinical and laboratory markers of disease progression in the three study patients in the 5 years before death. CD4 counts are shown as the mean counts over each year. Abbreviations: AZT: Zidovudine; PCP: Pneumocystis carinii pneumonia.
as an outgroup (Kumar et al., 1993). Phylogenetic analysis of the env region was confined to regions flanking the V1 and V2 hypervariable regions, because of the indeterminate and often arbitrary alignment of the hypervariable sequences and the possibility of positive selection. The number of synonymous substitutions at synonymous sites ($d_s$) and nonsynonymous substitutions at nonsynonymous sites ($d_n$) between sequences were estimated using the method of Nei and Gojobori (Nei et al., 1986). Finally, the bootstrap resampling method was used (500 replicates) to assess the confidence of each node in all trees constructed using the MEGA package.

3.4 STATISTICAL ANALYSIS.

The distribution of distances between and within patients were subjected to nonparametric statistical treatment using the Wilcoxon signed rank test included in the SYSTAT version 5.0 package.

3.5 RESULTS.

3.5.1 RATE OF SEQUENCE CHANGE IN $p17_\text{gag}$ REGIONS.

An unrooted neighbour-joining tree was constructed using 85 sequences from the $p17_\text{gag}$ region (positions 835-1270 in the HXB2 clone; Myers et al., 1991) from a range of lymphoid and non-lymphoid tissues of three HIV-infected
individuals dying in AIDS (p4, 5 and 6). Sequences from each of the three study patients were distinct, grouping separately into three clades. Bootstrap resampling supported the distinction of three separate groups (Fig 7).

All three study subjects were infected with HIV through drug abuse in 1982/83. Previous phylogenetic studies have implicated a common source of infection for the majority of drug users in Edinburgh, including the three described here (Holmes et al., 1995). The current sequence differences between the study subjects therefore must have originated from a process of divergent sequence change over a period of between 9 and 10 years. Using a mean figure of 9.5 years (or 19 years of divergent sequence change), the mean synonymous pairwise distances in the p17\(_{gag}\) region between individuals (0.149) indicated a rate of sequence change of 0.0077 substitutions per site per year. The rate of sequence change between pairs of individuals was similar, ranging from 0.006 to 0.009 (Table 6; Fig 8A-8C). This estimate was similar to those obtained in previous studies. For example, sequence comparisons in the p17\(_{gag}\) region of plasma RNA sequences from haemophiliacs infected from a common source indicated a mean rate of synonymous substitution in p17\(_{gag}\) of 0.006 to 0.0072 substitutions per site per year (Kasper et al., 1995).

The mean rate of nonsynonymous substitution between the study subjects was 0.058, lower than the silent rate. The mean \(d_n/d_s\) ratio of 0.39 indicated a bias towards silent substitutions in this region of the gag gene, consistent with previous estimates (Gojobori et al., 1990b; Myers et al., 1992b; Li et al., 1988b; Kasper et al., 1995b).
Fig. 7. Neighbor-joining tree of sequences in the p17 gag region of the three study subjects. Bootstrap values indicate the percentage of trees showing the observed patient-specific groupings.
### TABLE 6

SEQUENCE COMPARISONS BETWEEN STUDY SUBJECTS IN THE p17<sub> gag </sub> REGION

<table>
<thead>
<tr>
<th>Pair</th>
<th>Divergence (years)&lt;sup&gt;1&lt;/sup&gt;</th>
<th>No. pairwise comparisons</th>
<th>MEANS</th>
<th>Ratio (d&lt;sub&gt;s&lt;/sub&gt;/d&lt;sub&gt;o&lt;/sub&gt;)</th>
<th>Silent rate&lt;sup&gt;2&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>P4-P5</td>
<td>19</td>
<td>960</td>
<td>0.152</td>
<td>0.35</td>
<td>0.008</td>
</tr>
<tr>
<td>P4-P6</td>
<td>18</td>
<td>736</td>
<td>0.112</td>
<td>0.47</td>
<td>0.006</td>
</tr>
<tr>
<td>P5-P6</td>
<td>19</td>
<td>689</td>
<td>0.184</td>
<td>0.38</td>
<td>0.009</td>
</tr>
<tr>
<td>All</td>
<td>18.7</td>
<td>795</td>
<td>0.149</td>
<td>0.39</td>
<td>0.0077</td>
</tr>
</tbody>
</table>

<sup>1</sup>Based upon infection from a common source in 1982 (see Results)

<sup>2</sup>Silent substitution rate of sequence change between study subjects (per site per year)
Fig. 8. Frequency histograms of silent pairwise distances in the p17gog region between study subjects: (A) p4/p5; (B) p4/p6; (C) p5/p6, and within study subjects between sequences obtained from different tissues: (D) p4; (E) p5; (F) p6. Median values for distributions shown in Tables 6 and 7.
In this study I also determined the sequences of the V1 and V2 hypervariable regions and flanking regions in the env region from the three study patients (positions 6560-6876). Between individuals, the mean pairwise synonymous distance between sequences from the flanking regions (but omitting the hypervariable regions between 6623-6679 [V1] and 6701-6796 [V2]) was 0.104, lower than for the p17_gag region. In contrast the rate of nonsynonymous substitution in the V1/V2 flanking region was higher (0.083), producing an overall d_n/d_s ratio of 0.80, similar to previous estimates for the env region (Wolfs et al., 1990; Li et al., 1988).

In this study I used the measured rate of sequence change in the p17_gag region at silent sites to estimate the time of divergence between variants infecting different tissues within an infected individual. These data should indicate when the spread of HIV into non-lymphoid tissue occurred (Donaldson et al., 1994b).

3.5.2 PHYLOGENETIC ANALYSIS OF VARIANTS FROM DIFFERENT TISSUES.

Phylogenetic analysis was carried out using sequences from the p17_gag region and V1 and V2 flanking regions from a range of lymphoid and non-lymphoid tissues to determine the relatedness of variants between each tissue (e.g. lymph node, brain and lung; Fig 9). Bootstrap resampling using 500 replicate trees was carried out to estimate the robustness of the observed groupings.
Fig. 9. Phylogenetic analysis of sequences obtained from different tissues from the three study subjects (A, D: p4, B, E: p5; C, F: p6) in different subgenomic regions (A, B, C: p17gag region; D, E, F: V1/V2 flanking regions). Trees shown in rooted form using the unrelated subtype B sequence of HIV-MN as an outgroup. Bootstrap values of >75% indicated for branches highlighted in bold. Symbols: + Brain, ♦ Spinal cord, ● Lung ,■ Colon, □ Lymph node, ○ Spleen.
There was little evidence for consistent phylogenetic grouping by tissue origin. For example, p17 gag sequences from lymph node of patient 4 were found in two distinct lineages, one of which contained a variety of sequences from other tissues (lung and spleen; Fig 9A). Similarly sequences from brain were interspersed with those from colon, lung and spinal cord. In the V1/V2 flanking regions, a number of lineages were present supported by high bootstrap values (Fig 9D). For example, sequences from the brain were found in lineages a, b, c and f, all of which included sequences from various tissues such as brain, lymph node and lung. However, lineage a contains a large group of mainly brain isolates which may be an artifact of sampling error, whereby a very limited area of brain tissue has been sampled resulting in detection of a restricted population (see below).

Similar mixing of sequences from lymphoid and non-lymphoid tissues was observed amongst sequences from the other two study subjects (Fig 9B, 9C, 9E, 9F). For example, p17 gag sequences from both brain and lymphoid tissue of patient 5 were each found on lineages a and b, separated from each other by high bootstrap values (Fig 9B). In patient 6, sequences were obtained only from brain and lymph node but each of the lineages contained sequences from both sources (Fig 9C, 9F).
3.5.3 TIME OF DIVERGENCE OF HIV VARIANTS IN DIFFERENT TISSUES.

Pairwise synonymous distances between sequences from the p17 gag region from each patient were calculated to estimate the time of divergence of variants within each tissue. The previously established rate of sequence change in the p17 gag region of 0.0066 substitutions per site per year was used (Kasper et al., 1995), although similar results would have been obtained if I had used the synonymous substitution rate observed in this study (mean of three study individuals: 0.0077).

Mean synonymous pairwise sequence distances within study subjects were calculated by comparing sequences from all tissues with each other, as well as comparisons restricted to variants found in particular tissues, such as brain, lymph node and lung (Table 7; Fig 8D-8F). Comparison of variants found in all tissues produced a range of pairwise distances from 0.035 to 0.086, approximately a third of the mean inter-patient silent distance. These implied times of divergence of 2.6 to 6.5 years (Table 7).

In all three patients the mean distance between sequences from brain tissue was greater than the mean distance between variants in lymphoid tissue (Table 7; Fig 10), reflecting their wide distribution in multiple lineages by phylogenetic analysis (Fig 9). For example, the mean synonymous pairwise distances calculated for brain tissues ranged from 0.054 to 0.086 while that for lymphoid tissue ranged from 0.035 to 0.074 ($p<0.001$). These distances translate into average divergence
### Table 7: Sequence Comparisons of Variants from Different Tissues in the p17^R Region

<table>
<thead>
<tr>
<th>Tissue</th>
<th>No. of SEQS</th>
<th>Silent Sites</th>
<th>Non-Silent Sites</th>
<th>dN/dS</th>
<th>Mean p value</th>
<th>Significance of difference between pairwise distances amongst brain variants compared with other tissues</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brain</td>
<td>15</td>
<td>4</td>
<td>11</td>
<td>0.350</td>
<td>0.015</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Lymph Node</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>0.420</td>
<td>0.020</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Lung</td>
<td>11</td>
<td>7</td>
<td>4</td>
<td>0.350</td>
<td>0.015</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>All</strong></td>
<td>30</td>
<td>15</td>
<td>15</td>
<td>0.700</td>
<td>0.050</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brain</td>
<td>15</td>
<td>4</td>
<td>11</td>
<td>0.350</td>
<td>0.015</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Lymph Node</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>0.420</td>
<td>0.020</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Lung</td>
<td>11</td>
<td>7</td>
<td>4</td>
<td>0.350</td>
<td>0.015</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>All</strong></td>
<td>30</td>
<td>15</td>
<td>15</td>
<td>0.700</td>
<td>0.050</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brain</td>
<td>15</td>
<td>4</td>
<td>11</td>
<td>0.350</td>
<td>0.015</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Lymph Node</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>0.420</td>
<td>0.020</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Lung</td>
<td>11</td>
<td>7</td>
<td>4</td>
<td>0.350</td>
<td>0.015</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>All</strong></td>
<td>30</td>
<td>15</td>
<td>15</td>
<td>0.700</td>
<td>0.050</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

S/Including P4, 5 and 6 only.
Fig. 10. Distribution of pairwise distances in different tissues from the three study patients (p4, 5 and 6) at (A) silent sites and (B) non-silent sites in the pi7gag region.
times of 4.1 to 6.5 years and 2.65 to 5.6 years for the brain and lymphoid variants respectively. Overall, sequences between variants found in brain were no more similar to each other (0.080 for three study patients; Table 7) than they were to those present in lymphoid tissue (mean silent pairwise distance between brain and lymph node sequences: 0.070). In two of the patients mean synonymous pairwise distances were also calculated for variants isolated from lung tissue. In p4 the mean distance for lung variants (0.070) was greater than that for brain (0.054) and lymphoid variants (0.035), while in p5 the mean distance for lung variants (0.043) was approximately half that of brain (0.082) and very similar to lymphoid variants (0.048). In p5 focal inflammatory cell infiltrates were detected within lung tissue. Hence, sampling an area with such a focal infiltrate may provide an under-representation of the diversity of variants within this tissue. Indeed, the majority V3 sequence detected in lung tissue, lymphoid tissue and PBNCs was identical. In p4 patchy collections of macrophages and MGCs were present in the lung sections examined and the majority V3 sequence detected in lung tissue was identical to that detected in brain tissue, contrasting that found in p5.

Mean nonsynonymous pairwise distances were also calculated and were found to be lower than the distances at silent sites only. Nonsynonymous distances calculated for brain tissue only ranged from 0.010 to 0.042 and were higher than those observed between variants in lymphoid tissue of the three study individuals (0.012 to 0.020; p<0.001). Subsequently these values produced \( d_\text{N}/d_\text{s} \) ratios of between 0.158 and 0.49 for brain tissue only and 0.16 and 0.46 for lymphoid tissue only, similar to that observed previously using inter-patient comparisons. These
ratios indicate that most substitutions occurring within an individual in p17\_gag region were silent.

3.5.4 PHYLOGENETIC ANALYSIS OF SEQUENCE VARIANTS FROM PATIENT 79 ISOLATED FROM DIFFERENT TISSUES.

Following this study I examined the diversity of HIV-1 within a fourth individual (p79), from the Edinburgh IDU cohort, who had been infected for approximately 12 years to discern whether a similar distribution of viral variants from various tissues could be observed in an individual infected for a considerably longer period of time than p4, 5 and 6 (infected for 9 to 10 years). In this study subject I determined the sequences from the p17\_gag region only. Phylogenetic analysis was carried out to determine the relatedness of variants between each tissue (brain, lymph node and lung; Fig 11). Bootstrap resampling using 500 replicate trees was carried out to estimate the robustness of observed groupings. Consistent with the previous phylogenetic analyses, sequences from lymph node and lung tissues were interspersed throughout the tree, although fairly low bootstrap values were observed. Only two lineages had values higher than 75%; lineage a contained sequences from brain tissue only and lineage b contained sequences from lung tissue only. In contrast to the previous phylogenetic analysis, sequences obtained from brain tissue from this individual grouped closely together with only one variant falling outside the main group, being more closely related to a lymph node variant.
Fig. 11. Phylogenetic analysis of sequences obtained from different tissues of p79 in the p17 gag region. Trees shown in rooted form using the unrelated subtype B sequence of HIV-MN as an outgroup. Bootstrap values of >75% indicated for branches highlighted in bold. Symbols: + Brain, • Lung, □ Lymph node.
Pairwise synonymous distances between sequences from the p17<sub>gag</sub> region of p79 were calculated to estimate the time of divergence of variants within each tissue (Table 8; Fig 12). The previously established rate of sequence change for synonymous sites (0.0066 substitutions per site per year) was used (Kasper et al., 1995). In contrast with estimates from the previous study, the mean distance between sequences from brain tissue from p79 were considerably smaller than the mean distance between variants from lymphoid or lung tissue. For example, at synonymous sites the mean distance for brain tissue was 0.015, compared with 0.06 and 0.07 in lymph node and lung respectively. These distances translate into approximate mean divergence times of 1.14, 4.54 and 5.3 years for brain, lymph node and lung variants respectively. Mean synonymous pairwise distances between brain and lymph node and brain and lung sequences were calculated to be 0.042 and 0.044 respectively. Therefore, variants found in brain were far more similar to each other than they were to those found in lymphoid or lung tissue. Phylogenetic analysis reflects this limited diversification of brain variants with all brain sequences grouping tightly together.

Mean nonsynonymous pairwise distances were also calculated for P79 and were found to be lower than distances at silent sites only, consistent with the previous analysis (Table 8; Fig 12). These values were used to calculate d<sub>s</sub>/d<sub>s</sub> ratios for each tissue and produced ratios of 0.47, 0.4 and 0.44 for brain, lymphoid and lung tissues respectively. These ratios were similar to those observed previously for inter- and intra-patient comparisons and indicate that the majority of
<table>
<thead>
<tr>
<th>Patient Tissue</th>
<th>No. of SEQs</th>
<th>of Silent Sites</th>
<th>dN/dS</th>
<th>Mean p Value</th>
<th>Significance of difference amongst pairwise distances amongst brain variants compared with other tissues.</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>121</td>
<td>0.063</td>
<td>4.77</td>
<td>0.023</td>
<td>Mean p Value2</td>
</tr>
<tr>
<td>Brain</td>
<td>49</td>
<td>0.061</td>
<td>4.62</td>
<td>0.023</td>
<td>Mean p Value2</td>
</tr>
<tr>
<td>Lymph node</td>
<td>38</td>
<td>0.059</td>
<td>4.47</td>
<td>0.021</td>
<td>Mean p Value2</td>
</tr>
<tr>
<td>Lung</td>
<td>30</td>
<td>0.061</td>
<td>4.62</td>
<td>0.025</td>
<td>Mean p Value2</td>
</tr>
<tr>
<td>All</td>
<td>46</td>
<td>0.051</td>
<td>3.86</td>
<td>0.021</td>
<td>Mean p Value2</td>
</tr>
<tr>
<td>Brain</td>
<td>14</td>
<td>0.015</td>
<td>1.14</td>
<td>0.007</td>
<td>Mean p Value2</td>
</tr>
<tr>
<td>Lymph node</td>
<td>17</td>
<td>0.060</td>
<td>4.54</td>
<td>&lt;0.001</td>
<td>Mean p Value2</td>
</tr>
<tr>
<td>Lung</td>
<td>15</td>
<td>0.070</td>
<td>5.30</td>
<td>&lt;0.001</td>
<td>Mean p Value2</td>
</tr>
<tr>
<td>All</td>
<td>46</td>
<td>0.063</td>
<td>4.77</td>
<td>&lt;0.001</td>
<td>Mean p Value2</td>
</tr>
</tbody>
</table>

*Including P4, 5, 6 and 7.*
Fig. 12. Distribution of pairwise distances in different tissues from p79 at (A) silent sites and (B) non-silent sites in the p17_pro region.
substitutions which occurred within the p17_{pg} region were silent.

3.6 DISCUSSION.

3.6.1 DIVERSITY OF HIV-1 in vivo.

In this study a comparison of the diversity of HIV-1 in different tissues, including brain, lymph node and lung, obtained from four individuals infected from a common source in 1982/83 was carried out. Three of the individuals were found to have very diverse populations in all tissues although, in one individual (p79), relatively restricted diversity within brain tissue was observed. Similar results were obtained by Wong et al., who analysed the distribution of pol sequences within different tissue compartments from a number of patients who died in AIDS and who had received AZT treatment for variable periods of time (6 months to 2 years), prior to death. In three of the study subjects Wong et al., observed restricted diversity within brain tissue, while in a fourth subject a more diverse population was formed from two distinct evolutionary lineages, with one containing substitutions in pol at codons 41 and 215 that confer resistance to AZT treatment. Within the three individuals showing restricted diversity, differences in AZT resistance were noted. In subject A restricted diversity was seen in brain tissue and the entire population was found to be resistant to AZT with mutations at codons 41 and 215. Subject B and D also showed restricted diversity in brain however, in these two patients, no variants contained mutations associated with antiviral
resistance.

In this study two individuals (p4 and p79) underwent extensive AZT treatment for up to five years prior to death. Diverse populations were observed in the brain from p4 which contrasted to the very restricted diversity seen in the brain from p79. P6 was treated with AZT for two years prior to death and diverse populations were observed in brain. Finally p5, who remained untreated throughout infection, also showed diverse populations within brain tissue. Therefore, in this study and the study by Wong et al., both diverse and restricted populations have been detected in brain tissue from a number of individuals. This discrepancy cannot be attributed to AZT resistance since similar degrees of diversity have been observed in individuals with and without AZT resistant variants. For example, in subject A all variants detected were resistant to AZT yet this individual showed restricted diversity within the brain. Similarly, p5 had never been treated with any retroviral agents and yet displayed considerable diversity within brain tissue.

How can these differences in diversity within brain tissue be explained? One explanation could be that AZT resistant mutant variants may infect brain tissue and replace the initial infecting populations, thereby producing a homogenous brain population. This hypothesis would explain the restricted diversity in the pol region of subject A. However, complete replacement would be hampered by the lack of mobility of infected cells within a solid tissue. In the peripheral circulation, however constant viral replication and rapid cell turnover can lead to rapid population replacements within PBMCs (Coffin, 1996; Ho et al., 1995; Wei et al., 1995). Although we do not know the lifespan of HIV-infected
cells in the CNS it is less likely that such a process would occur so rapidly within brain tissue.

Another, more likely explanation for the observation of different populations within brain tissue could be due to inappropriate sampling. The brain itself is a large mass of solid tissue where cellular movement would be very restricted, unlike the peripheral circulation. It is therefore, highly likely that separate populations could emerge within the brain following initial infection. Brain tissue may be seeded early on in infection by infected cells from the peripheral circulation and form distinct foci of infection. Indeed, in this study I have shown that multiple lineages exist within brain tissue of three of the study subjects (p4, 5 and 6) and it is possible that each represents an independent focus of infection. Similarly in a previous study analysing the distribution of V3 sequences within different tissues, variants found in the left brain of one individual were not detected in the right brain of the same individual, again providing strong evidence to suggest multiple infection of brain tissue (Donaldson et al., 1994a). Following replication each focus of infection will have diverged from a single infected cell and therefore may display restricted diversity. The number of foci present will depend on how diffuse the initial seeding of the brain was.

When sampling brain tissue for analysis it is very possible that an unrepresentative population may be selected. For example, the tissue from which DNA was extracted may contain only one focus of infection, ultimately derived from a single infecting virus particle. Consequently, restricted diversity would be observed in any variants sampled from this tissue section and appear monophyletic.
by evolutionary analysis. Conversely, the tissue section analysed may contain a number of independent infected centres leading to a much more diverse, polyphyletic population. These differences could explain the variation observed between populations of HIV sequences in the brain tissue of p4, 5, 6 and 79. Similarly, in the study by Wong et al., the observation of two brain populations in subject C may have resulted from the presence of two foci of infection in the same tissue sample, one lineage being resistant to AZT while the other was sensitive. The monophyletic populations observed in the other three individuals may have resulted from sampling very small amounts of tissue which contained only one infection site. Unless the effect of sampling is taken into account, it is wrong to infer that the restricted diversity in one sample reflects that of the whole tissue. Conclusions regarding the biological relevance of such observations may be premature.

In the future studies should include multiple sampling sites to ensure that representative populations of HIV variants from the brain tissue are analysed. This theory could be investigated experimentally by dissection of foci of infection. For example, discrete foci of infection could be identified using in situ PCR, subsequently dissected from brain tissue and analysed in a similar manner to that described here. This type of analysis would discern whether each discrete focus of infection was indeed derived from a single infecting virus. From my own study, the observation that populations in the brain can be polyphyletic remains valid, however it is likely that the monophyletic populations observed in p79 and in the study by Wong et al., under-represent the true diversity of HIV-1 in the CNS.
3.6.2 RATE OF SEQUENCE CHANGE OF HIV in vivo.

In this study I have used published rates as well as estimates based upon the sequences recovered from three of the study patients (p4, 5 and 6), to estimate the times of divergence of variants infecting different tissues in vivo. Measurement of the rate of sequence change was possible for the study patients because it was known that all 3 patients were originally infected with HIV from a common source in an outbreak around 1982/1983, so that each was infected for approximately 9 to 10 years prior to death. Therefore between any 2 individuals, there was approximately 19 years of divergent sequence change. Synonymous rates of substitution were calculated for each patient in p17 gag (Table 6) and ranged from 0.006 to 0.009 substitutions per site per year (mean 0.0077), while the rate for the V1/V2 flanking regions was slightly lower (mean 0.0056 substitutions per site per year).

One assumption that must be made when calculating times of divergence from sequence distances is that the rate of sequence change remains constant throughout the course of infection, and there is little direct evidence that this is justified. Although higher levels of virus replication clearly occur later in the course of disease, this does not necessarily imply that the rate of sequence change should be higher. The rate of sequence change is proportional to the number of replication cycles, whose length is determined by the replicative processes within the cell, unless a substantial proportion of the sampled population originates from virus that has reactivated from latently infected cells where viral replication may
not have occurred for several years.

Empirically, however, the rate of sequence change at silent sites in \( p17_{gag} \) over the first 2 years of infection in haemophiliacs (0.0066 per site per year; Kasper et al., 1995) was similar to that observed in the three study patients (0.0077), in which the period of infection was 9-10 years, covering primary infection to death from AIDS. These figures are in turn within the range of those from several other studies using different observation periods and study subjects at different stages of disease (Gojobori et al., 1990b; Li et al., 1988b; Wolfs et al., 1990b).

Although the rate of non-silent sequence change in the \( gag \) region was lower than the synonymous rate, times of divergence based on non-silent sites provided similar times of divergence of variants in different tissues (mean time of divergence of variants within each subject: 3.1 years compared with 3.7 years using the published rate of silent sequence change in \( p17_{gag} \); Kasper et al., 1995). This is despite the theoretical possibility that the rate would be affected by phenotypic selection of variants with changes in the \( p17_{gag} \) region.

Using the mean synonymous rate of substitution for \( p17_{gag} \) of 0.0066 substitutions per site per year (Kasper et al., 1995), the average time of divergence between brain and lymph node variants within an individual patient were calculated (Table 7) and a range of values from 3.5 to 6.5 years was obtained. In lymphoid tissue, the mean diversity of \( gag \) sequences implied an approximate population age of 2.65 to 5.6 years, while those infecting brain were significantly more variable, suggesting an even earlier time of diversification (4.1 to 6.5 years). Despite the
large potential inaccuracies in calculating times of divergence based upon sequence distances, it is clear that compared with the total duration of infection within the patients (9 or 10 years), the observed diversity within brain tissue suggests infection occurred relatively early in the course of HIV infection, clearly preceding the onset of AIDS in two of the four study individuals (Fig 6).

In p79 the average time of divergence between brain and lymph node variants was calculated and found to be 3.2 years, within the range calculated for the other study patients. In lymphoid tissue an approximate population age of 4.5 years was calculated using the mean synonymous pairwise distance values, similar to the values calculated for the other three study subjects. However, variants infecting brain were considerably less variable with a diversification time of approximately 1.14 years. As previously suggested, this figure may underestimate the diversity of variants in the brain, through variants sampled from a restricted number of foci of infection. The similarity in the calculated divergence times of variants in lymphoid tissue and lung and those in the study by Wong et al., are consistent with the much greater degree of mixing of populations from trafficking of lymphocytes and macrophages through lymph nodes and other tissues.

3.6.3 ORGAN SPECIFIC DIFFERENCES OF HIV IN THE V3 REGION.

Populations of HIV variants infecting different tissues in vivo are generally distinct in the V3 hypervariable region of env (Ball et al., 1994; Korber et al., 1994; Power et al., 1994) including three of the patients in this study (Fig 13A-
For example, for p5, none of the V3 sequences of either the major population (15/17) or minor population (2/17) found in the brain were found amongst those from lung, PBMCs and lymph node; these latter tissues were dominated by a variant with a substitution at position 28 (35/42; Fig 13A). Similarly, for p6, variants in the brain were uniform, and differed from lymph node variants in all but one case by 1-3 amino acids (Fig 13B). The diversity of sequences in the V3 region of p4 made comparison more difficult, but again the main variants in brain (14/17) were not found in PBMCs or spleen (n=16) or, with a single exception, in lymph node (15 sequences; Fig 13C).

On the basis of this apparently tissue-specific distribution of variants in V3, it has been suggested that these population differences have adaptive significance and reflect different tropism for the different infected cell types in different tissues. The involvement of V3 would be consistent with the previous observation of its role in determining the ability of HIV to replicate in different cell types in vitro (Millich et al., 1993). Within macrophage-tropic isolates an acidic amino acid or alanine was predominantly seen at position 25 while a basic or uncharged amino acid at this position was associated with non-conservative basic amino acid substitutions at positions 11, 24 and 32 correlating with T-cell tropism, consistent with other studies (Hwang et al., 1991; Shioda et al., 1994; Morris et al., 1994; Kasper et al., 1994; Shioda et al., 1992; Westervelt et al., 1991; Chesebro et al., 1992). Extending this work, Power et al., compared cloned sequences from brain and spleen in demented and non-demented patients, and found evidence for specific amino acid substitutions at 2 positions in the V3 loop (histidine at position 305 and
FIG. 13. Proviral V3 loop amino acid sequences from infected organs from terminal AIDS patients (A) p4; (B) p5; (C) p6; (D) p9. All sequences compared to subtype B consensus sequence for V3 (LaRosa, 1990); ":": identity with subtype B consensus; "-": gap introduced to preserve alignment with consensus sequence; "n": number of sequences observed.
<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Consensus: CTRPNNNTRK
SIHIGPGRAF
YTTGEIIGDI
RQAHC

Brain
A

Brain (L)

Brain (R)

Colon

Lymph node

Lung

Spinal cord

Spleen
proline at position 329) that correlated with neurotropism and the clinical expression of HIV dementia (Power et al., 1994). However, while other studies have also found separate populations infecting brain compared with those infecting lymphoid tissue, there appears to be no conserved features of the V3 loop that correlate with neurotropism (Di Stefano et al., 1995; Korber et al., 1994; Reddy et al., 1996; Liu et al., 1990; Epstein et al., 1991; Keys et al., 1993).

Furthermore, there is no evidence for a correlation between tissue distribution with the predicted phenotype of such V3 sequences in vitro. For example, in a previous study of the three patients analysed in this study and others (Donaldson et al., 1994a), each tissue was found to be predominantly infected with variants with a predicted NSI /macrophage tropic phenotype, irrespective of tissue origin. In these cases, the observed amino acid differences between brain and lymphoid tissue were relatively few and probably unlikely on their own to alter the phenotype of the virus (Donaldson et al., 1994a; see below).

Other studies support the conclusion that the V3 region is to some extent involved in tissue tropism but that interaction with other regions in the HIV-1 genome is required for infectivity (Koito et al., 1994; Carrillo et al., 1996; Stamatakos et al., 1993; Carrillo et al., 1996). It has been suggested that mutations altering the structure of the V3 loop can affect the conformation of gp120 and that in turn the structure of the V3 loop is influenced by the conformation of other regions in gp120 (Stamatakos et al., 1993). An interaction of the V3 loop with a small region of the C4 domain has been suggested to be required for infectivity of Jurkat T cells lines (Carrillo et al., 1996a; Moore et al., 1993b; Wyatt et al.,
1992b; Morrison et al., 1993b). Therefore, although it is universally accepted that restricted variability exists in the V3 loop of HIV-1 gp120 there is no agreed interpretation of this observation.

3.6.4 MULTIPLE EVOLUTIONARY LINEAGES IN p17\textsubscript{gag} AND V1/V2 REGIONS.

Given the previously observed organ-specific populations in the V3 region, it was surprising to find a different relationship between variants when sequences elsewhere in the genome were compared. In both p17\textsubscript{gag} and V1/V2 flanking regions of p4, 5 and 6, I observed numerous independent lineages containing sequences from non-lymphoid tissues such as brain and lung mixed with those from lymphoid tissues. In p79, although a similar mixing of lymph node and lung variants to that seen in the other three study subjects was observed, brain variants all grouped within the same lineage distinct from the other tissues. Some of these groupings were confirmed by bootstrap resampling analysis (Fig 9 and 11). Comparison of the actual V1 and V2 sequences showed a pattern of sequence variability between tissues similar to that of the flanking regions, and without evidence of tissue specific groupings; this data is discussed in chapter 4.

There are at least three possible explanations for differences in grouping in different regions of the genomes; these include (a) different rates of sequence change in different tissues, (b) convergence and (c) recombination, and are reviewed below.
The first hypothesis, originally proposed by Korber et al. (Korber et al., 1994), is based upon the principle that infection of non-lymphoid tissue such as brain occurs early in the course of infection at a time when the viral population is relatively homogeneous in the V3 region. Therefore, variants infecting the brain would be initially similar to variants infecting non-lymphoid tissues. Subsequently, as disease progresses, variants found in lymphoid tissues may undergo more rapid sequence change in V3 and elsewhere in the genome associated with population replacements arising from immune escape or antiviral treatment. For example, variants resistant to neutralization or to antivirals such as AZT would outgrow other variants present within the lymphoid tissue. Rapid turnover and population replacements may be facilitated by the continuous movement of lymphocytes and other susceptible cells through lymphoid tissue. The previously estimated high rate of turnover of HIV-infected lymphocytes (Ho et al., 1995; Wei et al., 1995) following antiviral treatment is consistent with the existence of a relatively dynamic lymphoid-cell population, whereas at least for antiviral resistance, the brain population is not (Strappe et al., personal communication). Recently, evidence for extremely slow turnover of variants in brain tissue was obtained by sequence comparisons of the pol gene of variants infecting brain and lymphoid tissues from study subjects dying in AIDS (Wong et al., 1997). Despite frequently prolonged antiviral treatment prior to death, many individuals showed predominantly wild-type (i.e. AZT sensitive) variants in brain, while variants recovered from spleen and/or lymph node were predominantly or exclusively resistant.
Alternatively, variants infecting lymphoid cells may be subject to more rapid changes over time associated with changes in the V3 region that determine the shift in the phenotype of HIV upon disease progression. Variants in the brain however, may be unable to undergo such radical changes in the V3 region due to the continued strict requirements for replication in cells of the brain that are largely monocyte derived cells, i.e., infiltrating macrophages and microglia cells (Vazeux et al., 1987; Price et al., 1988). The survival of the original infecting population in the brain and its replacement in lymphoid cells would explain the former’s greater diversity in all parts of the genome other than those that determined tropism and the observed organ-specific differences in V3 populations. This hypothesis implies early entry of HIV into the brain, and although the V3 region is involved in tropism, it is not in the simple way it has been previously imagined.

It is possible to account for the organ-specific populations in V3 by other processes that do not necessarily imply early entry into the brain. For example, the organ-specific similarities in V3 sequences amongst variants that are not closely related in evolutionary terms could have originated from a process of strong convergent evolution, whereby the V3 sequence determines the ability of variants to grow in different cell types. Independent evidence for the existence of positive selection leading to convergence in V3 has been obtained from a study of haemophiliacs infected from a common source, who showed similarities in the pattern of sequence change in the V3 region in different individuals (Kasper et al., 1995). Similarly, a longitudinal study of a single infected individual showed several independent occurrences of certain amino acid changes in the V3 loop in variants
forming two evolutionarily distinct lineages (Holmes et al., 1992).

It is unlikely that the V3 loop could be the sole determinant of tropism, as the differences between populations infecting brain and lymphoid tissue are often trivial, and would be unlikely on their own to affect the phenotype of the virus. For example, all variants in the brain of p5 differed from those of lymphoid tissue at only one position (position 28), where a glutamate replaced an aspartate, a conservative amino acid change. Evidence for the functional equivalence of these two residues at this position can be inferred from their approximately equal distribution in isolates of the NSI phenotype, and amongst variants infecting a range of tissues collected at autopsy from these and other individuals (Di Stefano et al., 1995; Korber et al., 1994; Power et al., 1994; Reddy et al., 1996; Liu et al., 1990; Epstein et al., 1991; Keys et al., 1993). Furthermore, if convergence were the explanation for the organ-specific grouping of V3 sequences, we might expect to observe general similarities between variants infecting specific tissues from unrelated HIV-infected individuals. However, apart from the one study (Power et al., 1994), it has generally proved impossible to demonstrate any specific conserved sequence or motif in V3 or elsewhere in env that correlates with the cell type infected in vivo (see above). On the other hand, as noted above, it is possible the actual V3 sequence required for replication in different cell types may depend upon interactions between V3 and other regions of env, so that different V3 sequences may evolve to carry out equivalent functions in different HIV strains.

The other mechanism for different relationships in different parts of the genome is recombination, where a requirement for specific V3 sequences that
confers an ability to infect different tissues may favour recombination with an already divergent pre-existing population either within or out-with the tissue where the variants were found. Recombination occurs frequently in retroviruses including HIV-1, and is a mechanism by which genetic variation can be increased (Coffin, 1979). Recombination requires that multiple infection of cells occurs, and although there is evidence that this may occur in vitro (Rey et al., 1986; Koyanagi et al., 1987), the scarcity of HIV-infected cells in brain and other tissues seems to suggest that it may be an unlikely event in vivo. However, it is possible that recombinants may be generated elsewhere where high levels of replication occur (e.g. in lymphoid tissue) producing variants that are uniquely able to enter and replicate within the CNS.

Whether the similarities in V3 originated from convergence or recombination, these hypotheses suggest that the observed diversity of variants within brain tissue could have originated by a process of multiple entry from sources outside the CNS. Therefore, the actual duration of infection in the brain may be substantially lower than can be calculated by estimating its population diversity. Indeed, the grouping of variants from brain and lymphoid tissue by phylogenetic analysis of the p17gag and V1/V2 flanking regions could be regarded as evidence for a process of multiple entry. On the other hand this hypothesis does not easily explain how populations in brain should be consistently more diverse than those in lymphoid tissue or other presumed sources of infection in brain. The observed diversity of p17gag sequences in brain tissue, from patients 4, 5 and 6, is more consistent with the first hypothesis of a lower rate of population replacement
in brain compared with lymphoid tissue. In the case of p79 however, a less divergent population was detected in brain tissue (as previously discussed). Therefore, this finding may simply highlight the limitations encountered when sampling from solid tissues. In the future more rigorous sampling techniques should be applied to determine whether the difference in divergence observed in the four study patients here was due to insufficient sampling or not.

In summary, the main findings of this study were the observation of an unusually diverse population of HIV variants in brain of three of the study subjects, without evidence for any closer evolutionary relationship between them than to variants infecting other tissues in the body. Although late entry of recombinant viruses is a possibility, it is more likely that viral entry into the brain occurs relatively early in the course of disease, based upon observations of its higher diversity in brain than in other tissues and the existence of multiple evolutionary lineages containing sequences from brain. These findings suggest that loss of immune competence is not solely required for entry into non-lymphoid tissue, and the strong association between HIV-induced neuropathology and disease progression may be consequent to reactivation rather than de novo infection of the CNS. The finding of variants in the brain on several different evolutionary lineages challenges the hypothesis of the evolution of a uniquely neurotropic strain. It is possible that the only requirement for infection of the brain may be macrophage-tropism and hence the possession of a V3 loop sequence that is of low charge and shows few differences from the subtype B consensus sequence (Donaldson et al., 1994a).
This study represents the first attempt to use evolutionary analysis of variants infecting different tissues. The finding of different inter-relationships between variants in different parts of the genome, combined with uncertainty about the frequency and site of recombination in vivo and the selection pressures that could produce convergent evolution in V3, highlights the complexity in trying to understand the dynamics of HIV replication and dissemination to different tissues. However, this research at least provides a starting point for a more rigorous examination on the existence of HIV tropism in vivo.
CHAPTER 4: INVESTIGATION OF POPULATION DIVERSITY OF HIV-1

IN VIVO BY NUCLEOTIDE SEQUENCING AND LENGTH

POLYMORPHISM ANALYSIS OF THE V1/V2 HYPERVARIABLE REGION

OF env.
4.1 INTRODUCTION.

HIV-1 may exist in a state of permanent transition to survive external factors such as immune surveillance. This continual fluctuation has given rise to the term quasispecies to describe a population of closely related variants of HIV-1. Different viral variants, which exist within such a quasispecies, could potentially differ in their biological properties such as cellular tropism, cytopathicity, syncytium induction, replication rates and neutralization properties. A number of these differences in phenotype have been attributed to the envelope glycoprotein gp120. Previous studies have implicated the V3 domain, located within the third major cysteine loop, as the major determinant for these biological properties with specific amino acid changes in the V3 domain associated with differences in cell tropism and the ability to produce syncytia (Shioda et al., 1992; Shioda et al., 1991; Shioda et al., 1994; Chesebro et al., 1992; Millich et al., 1993; Morris et al., 1994; Power et al., 1994; Westervelt et al., 1991). This region has also been shown to be the principal target of neutralising antibodies produced by infection or immunization (Gorny et al., 1994a; LaRosa et al., 1990a; Freed et al., 1991a; Zwart et al., 1991a; Hwang et al., 1991a).

Functional determinants, however, are not solely restricted to the V3 domain of gp120. More recently, biological characteristics of HIV-1 have been attributed to the V1 and V2 hypervariable regions located in the second major cysteine loop (Boyd et al., 1993; Wu et al., 1995; Toohey et al., 1995; Sullivan et al., 1993; Koito et al., 1994; Westervelt et al., 1992; Koito et al., 1995). Sullivan
et al., have shown that the V1/V2 hypervariable regions may be involved in post-receptor binding events in the membrane fusion process (Sullivan et al., 1993). They carried out a mutational analysis showing that the replacement of certain highly conserved amino acid residues resulted in envelope glycoproteins deficient in syncytium formation and/or virus infectivity. Similarly, Wang et al., using site directed mutagenesis, identified a conserved aspartic acid residue required in the early stages of HIV-1 replication (Wang et al., 1995). Replacement of this residue with either an alanine, leucine or glutamic acid residue resulted in a marked reduction in growth kinetics when compared with the wild-type phenotype. It is therefore conceivable that this aspartic acid residue is critical for postreceptor binding events and may interact with other regions in the gp120 molecule to confer the correct conformation required for virus entry.

Several structural features within the V2 region have been suggested to determine the phenotype of the virus (Groenink et al., 1993; Fouchier et al., 1995; Andeweg et al., 1995). Andeweg et al., constructed chimaeric envelope genes to delineate the regions of the envelope glycoprotein important in membrane fusion. They reported that in SI variants, the N-terminal part of the V2 region was likely to form a stable α-helix not predicted for NSI variants (Andeweg et al., 1995). Groenink et al., observed a greater length and overall positive charge of the V2 domain amongst SI, non-macrophage tropic isolates (Groenink et al., 1993). However, in a follow up study, only 2 out of 11 study patients were found to have elongated V2 domains 11 to 60 months prior to SI conversion (Fouchier et al., 1995). In an extension of this analysis, variants infecting 11 out of 12 individuals
(3 months after SI conversion) and 14 out of 26 individuals (6 months after SI conversion) were shown to have elongated V2 domains. These findings suggest that elongated V2 domains may be required transiently for SI conversion but are not necessarily required for maintenance of this phenotype (Schuitemaker et al., 1995; Fouchier et al., 1995). Subsequent investigations have effectively discounted a causal association between phenotype and primary amino acid sequence of V2 (Cornelissen et al., 1995; Palmer et al., 1996; Wang et al., 1995), even though determinants of cytopathology and cellular tropism are clearly in part dependent on this and neighbouring regions of gp120.

The V2 domain has also been shown to determine efficient infection of macrophages (O’Brien et al., 1990; Westervelt et al., 1992; Toohey et al., 1995; Koito et al., 1994; Boyd et al., 1993; Koito et al., 1995). Koito et al., generated T cell tropic/macrophage tropic recombinant viruses to investigate whether the V1/V2 domains were required for macrophage tropism (Koito et al., 1994). The results obtained strongly indicated a role for the V1/V2 regions in macrophage tropism and suggested these regions may interact with the V3 region modulating the overall conformation of gp120. Toohey et al., suggested that without the appropriate V1/V2 sequence macrophage tropic clones failed to spread following initial infection again suggesting a role for the V1/V2 regions in membrane fusion (Toohey et al., 1995).

The lack of correlation between viral phenotype and primary amino acid sequence of V1 and V2 can be clearly seen in Fig 14. Published V1 and V2 sequences of known phenotype from 3 independent studies (Groenink et al., 1993;
FIG 14. Comparison of the relationship between charge, length and glycosylation of V1 and V2 domains with the experimentally determined \textit{in vitro} phenotype of previously described HIV isolates. (A), (B): charge and length for V1 and V2 domain respectively; symbols: "\(\bigcirc\)" NSI isolates; "\(\bullet\)" SI isolates. (C), (D): number of potential N-linked glycosylation sites for V1 and V2 regions respectively. (see Appendix V).
Wong et al., 1995; Cornelissen et al., 1995), were analysed collectively and the overall charge and length were compared (see Appendix V). There was considerable variability in the overall charge of both V1 (-5 to +2) and V2 (-3 to +4) and in the length (V1: 16 to 49 amino acids; V2: 38 to 61 amino acids). They also differed in the position and number of several of the potential N-linked glycosylation sites, although some were highly conserved, as were the cysteine residues that maintain the structure of the V1 and V2 loops (Leonard et al., 1990). However, there was no systematic difference in these properties between NS1, macrophage tropic isolates and SI variants in either of these regions.

It has also been shown that V1 and V2 regions of gp120 can act as targets for neutralizing antibodies (Warrier et al., 1994; van Tijn et al., 1989; Ho et al., 1991; Yoshiyama et al., 1994; Moore et al., 1993; Gorny et al., 1994; Fung et al., 1992; Sullivan et al., 1993). Monoclonal antibodies with virus neutralizing activity have been mapped to the V2 region by binding to peptides corresponding to V2 sequences (Fung et al., 1992; McKeating et al., 1993), mutational analysis (Sullivan et al., 1993) and by competition with previously mapped anti-V2 monoclonal antibodies (Jeang et al., 1993). Consequently, linear, conformational and glycan dependent epitopes have been detected.

Finally, it has been suggested that the existence of functional interactions between V1/V2 hypervariable regions and other regions of gp120 are essential for viral infectivity and syncytium induction (Cho et al., 1996; Willey et al., 1989; Andeweg et al., 1993; Freed et al., 1994). Willey et al., have previously shown that substitution of an asparagine with a glutamine in the C2 region of gp120
generated a non-infectious virus (Willey et al., 1988), however revertant virions were found to emerge during long term coculture. Sequence analysis of these revertants revealed the substitution of an asparagine for a serine in the V2 region of gp120 compensating for the substitution in the C2 region and restoring virion infectivity (Willey et al., 1989). In another study, substitution of a tyrosine by a histidine at position 435 in the C4 region of gp120 prevented the binding of two conformational dependent anti-V1/V2 monoclonal antibodies. However, both monoclonal antibodies were able to bind the V1/V2 region in the absence of C4 suggesting that this region may affect the structure of the V1/V2 region (McKeating et al., 1993). Wang et al., have shown that substitution of a valine with an isoleucine at position 84 in C1 can compensate for amino acid substitutions in the V1/V2 regions which impair virus infectivity, demonstrating that the V1/V2 regions can functionally interact with C1 also (Wang et al., 1996).

Compared with the extensive analysis of V3 genetic variability in vivo and in vitro, there are few corresponding studies of the V1 and V2 regions. Analysis of the V1/V2 regions is more arduous since both these regions display extensive length variation as a result of insertions, deletions and duplications. This extensive length variation can occur within epidemiologically related groups and indeed within infected individuals making alignment of sequences far more difficult. In this study, I have used nucleotide sequencing and length polymorphism analysis (LPA) to investigate correlations between variability of V1 and V2 regions with disease stage and tissue origin. Tissues from various lymphoid and non-lymphoid organs were obtained at autopsy from a number of HIV-1 positive individuals.
Phylogenetic analysis of V1 and V2 hypervariable regions was carried out in order to explore the \textit{in vivo} distribution of the various lineages of HIV-1 present in each tissue. I also investigated whether any specific sequence characteristics existed which could differentiate between individuals at different stages of disease, or between different infected tissues.

\section*{4.2 STATISTICAL ANALYSIS.}

Sequence comparisons between viruses from three of the study subjects (P4, 5 and 6) were made in the V1 and V2 hypervariable regions. The V1 and V2 region amplified began at nucleotide 6539 of HXB2 and extended to position 6976. The length of the V1 and V2 regions used for sequence comparisons were 142 and 193 nucleotides respectively. Unrooted phylogenetic trees based upon uncorrected pairwise distances between nucleotide sequences obtained from lymph node, lung and brain samples from study subjects 4, 5 and 6 were constructed by a combination of the MEGA package (Kumar \textit{et al}., 1993) and the PHYLIP package, using the programs SEQBOOT, NEIGHBOR and DRAWTREE (version 3.5; Felsenstein, 1989). Statistical analysis was performed with the Mann-Whitney U test and data considered statistically significantly different when $p < 0.05$ (SYSTAT version 5.0 package).
4.3 RESULTS.

4.3.1 INTRA-SAMPLE VARIATION IN V1 AND V2 HYPERVARIABLE REGIONS.

I compared the 87 V1 and V2 amino acid sequences from the three HIV infected individuals who died in AIDS (Fig 15a-15c), and found a high degree of variability in both of these regions, although it was less pronounced in the V2 region. Four cysteine residues defined a double loop structure in V1 and V2 (Leonard et al., 1990) and were uniformly conserved in all sequences (residues 10, 15, 56 and 108). Considerable diversity was observed in the V1 hypervariable region both within and between individuals. The V2 hypervariable region, although showing a great deal of interpatient variation, was less variable than the V1 region within individuals. Few identical amino acid sequences were isolated from the same individual. For example, in p4 a total of 37 sequences (brain-12, lymph node-10 and lung-15) were isolated and 25 of these sequences were distinct (brain-5/12, lymph node-10/10 and lung-10/15). The majority of sequence variability was located between residues 82 and 103. In the more conserved regions of the V2 hypervariable loop (residues 56 to 81 and 104 to 108) there appeared to be a bias towards the conservation of charged amino acids. For example residues 65, 66, 67, 70, 71, 77 and 79 were all well conserved both between and within individuals. However, the V1 hypervariable region showed a number of substitutions of charged
FIG 15A. Proviral V1 and V2 domain amino acid sequences from p4. All sequences are compared with HIV$_{MN}$ sequence for V1 and V2. Symbols: ":": identity with HIV$_{MN}$; ":-": gap introduced to preserve sequence alignment.
FIG 15B. Proviral V1 and V2 domain amino acid sequences from p5. All sequences are compared with HIV_{MN} sequence for V1 and V2. Symbols: ".": identity with HIV_{MN}; "-": gap introduced to preserve sequence alignment.
FIG 15C. Proviral V1 and V2 domain amino acid sequences from p6. All sequences are compared with HIV$_{MN}$ sequence for V1 and V2. Symbols: ".": identity with HIV$_{MN}$; ":": gap introduced to preserve sequence alignment.
amino acids. There were however three charged amino acid residues that were well conserved in the V1 region (residues 17, 52 and 54) which were located in close proximity to the two cysteine residues.

Both hypervariable regions had several potential N-linked glycosylation sites (N-X-S/T; N= asparagine, X= any amino acid except proline, S= serine, T= threonine) which were well conserved throughout all the sequences analysed (residues 14, 55, 59 and 109). These were located in close proximity to the cysteine residues which define both loops and therefore may be involved in maintaining the conformation of these two hypervariable regions. The V1 hypervariable region also contained a number of more variable N-linked glycosylation sites brought about by amino acid substitutions and insertions. Additional N-linked glycosylation sites were observed in some V2 sequences from insertions in the more variable region (residues 82 to 103). In addition, the V1 region contained a number of serine/threonine rich insertions which may lead to the addition of O-linked carbohydrates (see discussion).

4.3.2 V1 AND V2 SEQUENCE VARIABILITY AND TISSUE TROPISM.

I compared the overall charge (Fig 16A-16F), length of variants (Fig 17A-17F) and number of potential glycosylation sites (Fig 18A-18F) of V1 and V2 sequences obtained from the three study subjects (see Appendix VI). For p4 and p5, there was some evidence for differences in charge between variants from brain compared with lymph node or lung (Fig 16A-16F). In p4 and p5, the calculated
FIG 16. Comparison of overall charge of V1 and V2 sequence variants from p4 (A, D), p5 (B, E) and p6 (C, F) in V1 (A, B, C) and V2 (D, E, F) regions. Significantly different distributions ($p<0.05$ using Mann-Whitney U test for non-normally distributed data) indicated by horizontal bar. BR, brain; LG, lung; LN, lymph node. (see Appendix VI).
FIG 17. Comparison of overall length of V1 and V2 sequence variants from p4 (A, D), p5 (B, E) and p6 (C, F) in V1 (A, B, C) and V2 (D, E, F) regions.

Significantly different distributions ($p<0.05$ using Mann-Whitney U test for non-normally distributed data) indicated by horizontal bar. BR, brain; LG, lung; LN, lymph node.(see Appendix VI).
FIG 18. Comparison of the number of potential N-linked glycosylation sites in V1 and V2 sequence variants from p4 (A, D), p5 (B, E) and p6 (C, F) in V1 (A, B, C) and V2 (D, E, F) regions. Significantly different distributions (p<0.05 using Mann-Whitney U test for non-normally distributed data) indicated by horizontal bar. BR, brain; LG, lung; LN, lymph node. (see Appendix VI).
overall charge of V2 of variants from the brain were significantly higher than those from the lung \((p=0.055\) and 0.014). In contrast, the V1 sequences from brain of p4 showed a significantly lower charge compared to variants from lymph node and lung (both \(p=0.001\)). However, this difference was not found in the other two study subjects.

No significant difference between the length of the V1 and V2 regions with tissue origin was observed for patient 4 and 6. However, in patient 5, a significant difference between brain and lung variants was observed in both V1 and V2 regions \((p=0.006\) and <0.001 respectively). Also, in the V2 region, lymph node variants from patient 5 were significantly longer than those from brain tissue \((p=0.008;\ Fig\ 17A-17F)\).

The number of potential N-linked glycosylation sites were also compared in both V1 and V2 regions between tissues from each patient \((Fig\ 18A-18F)\). In patients 4 and 5 a significant difference in the number of glycosylation sites in brain and lung tissues was apparent in both V1 and V2 regions. In the V1 region from patient 4 brain variants were found to have significantly fewer glycosylation sites when compared with lung variants \((p=<0.001)\), while in the V2 region lung variants were found to have significantly fewer glycosylation sites \((p=0.03)\).

Similarly, in patient 5, brain variants in the V1 and V2 regions were found to have significantly fewer glycosylation sites than lung variants \((p=0.017\) and <0.001). In only one patient was there a significant difference in the number of glycosylation sites between brain and lymphoid variants. In the V2 region of patient 5 brain variants were found to have significantly fewer glycosylation sites when compared
with that of lymphoid tissues \( p=0.001 \).

The significance of these differences is difficult to interpret as the distribution of values compared were derived from populations that were in some cases closely related genetically, therefore do not constitute independent observations. However, from this analysis, there were no obvious features from the primary sequences of V1 or V2 that correlated with tissue origin.

In the absence of any reproducible specific amino acid sequence differences between tissues, I calculated the overall degree of divergence between sequences from each individual, and used these uncorrected pairwise distances to construct unrooted neighbour-joining trees. The degree to which sequences group together in the tree was proportional to their overall similarity, while bootstrap resampling indicated the robustness of the observed groupings (Fig 19A-19C). Although there were differences between tissues in the frequencies of certain amino acids at particular sites, none were clearly associated with specific tissues. For example, in patient 4 at position 24 of the V1 domain, the majority of brain-derived sequences (8 of 11) had an aspartic acid (D) at this position, while this amino acid was absent in the majority of sequences obtained from lymph node (1 of 10) and lung (0 of 15). Similarly, at positions 65 and 83 of the V2 domain, the majority of brain-derived sequences had an asparagine (N) and valine (V) at these positions respectively. However, again these amino acids were absent in the majority of sequences derived from lymph node (position 65: 0 of 10; position 83: 0 of 10) and lung (position 65: 2 of 15; position 83: 7 of 15). In brain-derived sequences, from patient 5, the majority had a lysine (K) at position 20 (7 of 11), glutamine
FIG 19A. Neighbour-joining trees of sequences in V1 and V2 domains from patient 4. Symbols: Brain ▲; Lymph node □; Lung ●. Bootstrap values indicate the percentage of trees showing the observed phylogenetic groupings.
FIG 19B. Neighbour-joining trees of sequences in V1 and V2 domains from patient 5. Symbols: Brain △; Lymph node □; Lung ○. Bootstrap values indicate the percentage of trees showing the observed phylogenetic groupings.
FIG 19C. Neighbour-joining trees of sequences in V1 and V2 domains from patient 6. Symbols: Brain ▲; Lymph node □. Bootstrap values indicate the percentage of trees showing the observed phylogenetic groupings.
(Q) at position 82 (10 of 11), glycine (G) at position 84 (9 of 11), arginine (R) at position 102 (7 of 11) and asparagine (N) at position 107 (9 of 11). However, the corresponding amino acids at these positions were absent in all of the lymph node-derived sequences and present as minor populations in lung-derived sequences at positions 20 (1 of 10), 82 (3 of 10) and 102 (1 of 10) only. In patient 6 there were no discernable amino acids present in the majority of brain-derived sequences when compared with those from lymph node tissue.

These differences were reflected in the phylogenetic trees constructed for each patient. For example, sequences from brain of p4 were found in all four lineages separated by high bootstrap values (Fig 19A), while, lineages a, b and d also contained a number of sequences from lymph node and lung tissue. In the other two study subjects a similar mixing of variants from lymphoid and non-lymphoid tissue was observed. For example, in p5 five distinct lineages were observed separated by high bootstrap values, three of which contained sequences from lymph node and lung tissue (Fig 19B). Two lineages contained brain-derived sequences only (lineages a and e), reflecting the differences observed in the amino acid sequences (Fig 15A). In p6, each of the three lineages contained sequences obtained from brain and lymph node tissue (Fig 19C).

4.3.3 ANALYSIS OF NUMBER OF V1/V2 LENGTH VARIANTS AND DISEASE STATUS.

To investigate the accuracy of length polymorphism analysis (LPA) and its
suitability for population analysis, I compared length profiles of V1 and V2 with the range of predicted lengths derived from individual sequences obtained from the three study subjects (p4-p6). 1 µg of DNA was amplified, using nested PCR, from lymph node and brain tissue from each individual. This amplification was carried out in triplicate to demonstrate that representative populations were compared (Fig 20A-20B). A good concordance was observed between the number and length of variants obtained using LPA with the actual lengths of variants obtained from sequencing (Table 9a-9b), although LPA also detected minor variants not represented among the nucleotide sequences. From this initial investigation I found no consistent differences between lymph node and brain tissue. For example, in p4 the same length variant was observed in both tissues in V1 and V2. However, in patients 5 and 6 a number of different length variants were observed in both tissues.

Subsequently this technique was applied to examine the diversity and overall length of V1 and V2 variants infecting brain and lymphoid tissues of a larger study group (see Appendix VII and VIII). This comprised samples from 8 pre-symptomatic and 34 symptomatic individuals at time of death. Length analysis of variants from brain tissue was confined to patients with evidence of giant cell encephalitis (GCE) since low levels or undetectable frequencies of proviral sequences were observed in individuals without GCE (Bell et al., 1996; Donaldson et al., 1994). The number and length of variants from non-lymphoid and lymphoid tissue differed considerably between study subjects (Fig 21). I compared the number of different length variants obtained from tissues from pre-symptomatic
FIG 20A. LPA of HIV variants obtained from lymph node and brain tissue of p4, p5 and p6 for the V1 region. Sizes of bands in amino acids indicated.
FIG 20B. LPA of HIV variants obtained from lymph node and brain tissue of p4, p5 and p6 for the V2 region. Sizes of bands in amino acids indicated.
### TABLE 9

<table>
<thead>
<tr>
<th>Subject</th>
<th>Organ</th>
<th>No.</th>
<th>Observed Lengths by LPA</th>
<th>+</th>
<th>Actual Lengths</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>27 30 31 32 33 35 36 38 40 41 42</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>36</td>
<td>27 30 31 32 33 35 36 38 40 41 42</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>38</td>
<td>27 30 31 32 33 35 36 38 40 41 42</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>40</td>
<td>27 30 31 32 33 35 36 38 40 41 42</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>41</td>
<td>27 30 31 32 33 35 36 38 40 41 42</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>42</td>
<td>27 30 31 32 33 35 36 38 40 41 42</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Comparision of Lengths of V1 and V2 Obtained by LPA compared with Direct Sequencing**
<table>
<thead>
<tr>
<th>Observed lengths</th>
<th>No.</th>
<th>Subject</th>
<th>Organ</th>
<th>LNBR</th>
</tr>
</thead>
<tbody>
<tr>
<td>38 40 41 43 44 47 48</td>
<td>2 10</td>
<td>V2 Region</td>
<td>BR</td>
<td>+</td>
</tr>
<tr>
<td>38 40 41 43 44 47 48</td>
<td>2 10</td>
<td>V2 Region</td>
<td>LN</td>
<td>+</td>
</tr>
<tr>
<td>38 40 41 43 44 47 48</td>
<td>3 9 1</td>
<td>V2 Region</td>
<td>BR</td>
<td>+</td>
</tr>
<tr>
<td>38 40 41 43 44 47 48</td>
<td>2 10</td>
<td>V2 Region</td>
<td>LN</td>
<td>+</td>
</tr>
<tr>
<td>38 40 41 43 44 47 48</td>
<td>1 12</td>
<td>V2 Region</td>
<td>LN</td>
<td>+</td>
</tr>
</tbody>
</table>

The heavy cross (+) indicates more prominent bands on LPA (see Fig. 5).

Figures refer to the number of sequences obtained for each length.

Total number of different length variants detected by LPA or by direct sequencing.
FIG 21. Plot of CD4 lymphocyte count -v- population diversity as determined by LPA in V1 (A) and V2 (B) regions; Symbols: "+" Brain from study subjects with GCE; "●": lymph node from study subjects with AIDS; "○": lymph node from pre-symptomatic individuals at times of death. (C, D): Comparison of the distribution of length variants observed in brain from study subjects with GCE (BR), lymph nodes from study subjects with AIDS (LN) and from pre-symptomatic individuals at time of death (P) for V1 (C) and V2 (D). (see Appendix VII).
and symptomatic individuals with CD4 count and found no significant difference between each group in either the V1 or V2 region (Fig 21A-21B). Similarly there was no correlation between the length of variants with disease status (Fig 21C-21D). However, the V2 region was on average longer amongst variants from brain compared to those from lymph node from symptomatic and pre-symptomatic study subjects, although their ranges overlapped considerably ($p=0.028$ and 0.043 respectively). Furthermore, variants from brain showed a significantly lower number of length variants (diversity) than those from lymph node of symptomatic patients in the V2 region (mean number of length variants in brain: 1.3 and lymph node: 2.5; $p=0.00$), and pre-symptomatic patients in both V1 and V2 (mean number of length variants, V1: 4.1; V2: 3; $p=0.039$ and 0.00 respectively; Fig 21A-21B).

4.4 DISCUSSION.

4.4.1 LACK OF TISSUE SPECIFIC GROUPING BY SEQUENCE VARIATIONS IN V1 AND V2 HYPERVARIABLE DOMAINS.

This study was carried out to analyse tissue distribution and interpatient variability of V1 and V2 hypervariable domains of HIV-1. A number of V1 and V2 sequence variants from lymphoid and non-lymphoid tissues were analyzed from three HIV infected individuals dying in AIDS. Examination of primary sequences revealed no specific amino acid motifs in either the V1 or V2 regions that
correlated with tissue origin. A great deal of length diversity was observed within the V1 and V2 regions associated with either the addition or removal of potential N-linked glycosylation sites.

The V1 and V2 regions may be determinants of HIV-1 tropism and a number of studies have suggested that these two regions may be involved in determining the phenotype of the virus. However, there is little evidence to support the antigenicity of the V1 region. Van Tijn et al., (1989) reported the detection of antibodies to a V1 peptide, corresponding to residues 19 to 34 in our numbering system, in 8 acutely infected individuals. From the three patients analysed here a number of amino acid changes can be seen within this region, although the vast majority of sequences contain 2 or 3 glycosylation sites which may effect envelope conformation. Indeed, Gram et al., recently reported that the lack of an N-linked glycosylation site in the V1 loop rendered the mutant virus less sensitive to neutralization by V3 monoclonal antibodies and soluble CD4, suggesting the degree of glycosylation in the V1 region may modulate the tertiary structure of gp120 (Gram et al., 1994). Moore et al., have described a strong linear epitope, between residues 40 and 51 in our numbering system, recognized by antibodies from the serum of an infected lab worker (Moore et al., 1993a). However, this region is highly variable in the three patients examined here and may therefore be highly type specific.

A number of neutralization epitopes (linear, conformation dependent and glycan dependent) have been identified in the V2 region (Jeang et al., 1993; Sullivan et al., 1993; Fung et al., 1992; McKeating et al., 1993). Warrier et al.,
described a novel glycan-dependent epitope in the V2 region encompassing residues 60 to 67 in our numbering system (Warrier et al., 1994). Similarly, McKeating et al., described a linear epitope within this region of V2 (residues 60 to 70 in our numbering system; McKeating et al., 1993). This region is relatively well conserved within each of the patients described in this study. In patient 4, however an asparagine (N) at position 65 is prominent in brain isolates (10 of 12) but rare in lung (2 of 15) and absent in lymph node isolates (0 of 10). This amino acid was reported to have functional relevance in the C108G epitope described by Warrier et al., (Warrier et al., 1994), suggesting this region may be of importance during infection. McKeating et al., have also described a number of conformational epitopes located in the carboxy-terminal of the V2 loop (McKeating et al., 1993). The carboxy-terminal from the 3 patients studied here show some degree of variability, although there was a preponderance towards the presence of a charged residue at position 87. Patients 4 and 6 contained an N-linked glycosylation site within this region which is conserved throughout all the sequences in these two patients. Similarly, patient 5 was highly glycosylated within this region (1 to 4 CHO sites). These features may therefore contribute to the maintenance of antigenic epitopes. Indeed if these epitopes are well exposed on the virion surface, as proposed, extensive glycosylation may help to mask these regions facilitating escape from the immune response.

Carbohydrate residues have previously been implicated in the adhesion of the gp120 molecule to CD4 (Matthews et al., 1987). Using enzymatic deglycosylation Bernstein et al. (Bernstein et al., 1994) found evidence for the
modification of gp120 by O-linked carbohydrates in addition to N-linked carbohydrates. Unlike N-linked glycosylation there is no defined sequence for O-linked glycosylation and the sites for carbohydrate addition were not experimentally determined. However, studies of simian immunodeficiency virus (SIV) have documented extended sequences, in the region homologous to V1 in gp120 of HIV-1, rich in serine and threonine residues (TTTSTTTT), that resemble known O-linked glycosylation sites in other proteins (Jentoft, 1990; Overbaugh et al., 1992). In this study serine/threonine rich insertions were observed in the V1 region which, although differing in sequence to those found in variants of SIV, may lead to the addition of O-linked as well as N-linked carbohydrates. This may be an important factor for the infectivity of HIV-1 as a number of diverse functions have been attributed to carbohydrate post translational modifications including cell recognition and cell adhesion (Paulson, 1989; Brandley et al., 1986), escape from immunological constraints (Ezekowitz et al., 1989) and reduced ability to bind antibody (Davis et al., 1990). Syncytium formation by HIV-1 has also been shown to be blocked by antibodies to O-linked carbohydrate structures (Hansen et al., 1991). Therefore, glycosylation of hypervariable regions may play an important role in the infectivity and pathogenicity of HIV-1.

4.4.2 RELATIONSHIP BETWEEN V1 AND V2 SEQUENCES AND TISSUE DISTRIBUTION.

To examine the tissue distribution of these variants unrooted neighbour-
joining trees were constructed. Numerous independent lineages were observed containing sequences from non-lymphoid tissues such as brain and lung mixed with those from lymphoid tissues. A number of these groupings were confirmed by bootstrap resampling analysis (Fig 19A-19C). This lack of tissue specific groupings of V1/V2 domains was consistent with a previous study of the evolutionary analysis and tissue distribution of the \textit{p17gag} region and flanking regions of V1 and V2 from the same three individuals (Hughes \textit{et al.}, 1997). In this previous study I found multiple evolutionary lineages in both of these regions of the HIV-1 genome. Using the \textit{p17gag} region the time of diversification of \textit{in vivo} variants from brain tissue was estimated at 4.1 to 6.2 years, suggesting infection of brain tissue may occur as an early event in disease preceding the onset of AIDS. The lack of organ specific groupings contrasts with previous comparisons in the V3 region in both my own and other investigations (Niedrig \textit{et al.}, 1994; Reddy \textit{et al.}, 1996; Ball \textit{et al.}, 1994; Power \textit{et al.}, 1994; Epstein \textit{et al.}, 1991; Donaldson \textit{et al.}, 1994) and may reflect different rates of sequence turnover in different tissues, combined with different constraints on the sequence of V3 for infectivity in different cell types (see chapter 3). This explanation would suggest that due to the lack of segregation of V1 and V2 sequences observed in this study, these two regions of the HIV-1 genome do not influence the tropism of HIV-1 to the same extent as V3 or do so in such a way that is not apparent from comparisons of primary amino acid sequences (see below).

Many phenotypic differences between isolates of HIV-1, such as syncytium induction have been mapped to V3, where it has been shown that SI variants
generally have a higher overall charge and a greater number of amino acid differences from the consensus subtype B sequence (Millich et al., 1993). However, there is currently a consensus view that V2 (and V1) sequences of NSI and SI variants do not consistently differ from each other in overall charge, length or number of potential glycosylation sites (Fouchier et al., 1995; Palmer et al., 1996; Wang et al., 1995; Cornelissen et al., 1995). In this study, a similar lack of correlation was found between charge, length and number of potential glycosylation sites amongst variants amplified from different tissues. There was no evidence for any specific amino acid motif that correlated with tissue origin in either the V1 or V2 domains, nor a difference in length of V2 between pre-symptomatic and AIDS study subjects, despite the greater frequency of isolation of SI variants from the latter group.

These findings suggest that both V1 and V2 may be irrelevant to tissue tropism of HIV-1 or may contribute in a more subtle way undetectable by examination of primary sequences. For example, several studies have shown that V1 and V2 domains may co-operate with other regions of the envelope protein in determining cellular tropism of HIV-1 (McKeating et al., 1993; Sullivan et al., 1993; Koito et al., 1995; Wyatt et al., 1995; Andeweg et al., 1993; Freed et al., 1994; Koito et al., 1994; Carrillo et al., 1996; Groenink et al., 1993). It seems unlikely that genomic regions which have been shown to influence viral infectivity and post binding events should not somehow influence cytopathology and tissue tropism. However, the evident degree of sequence flexibility tolerated by the virus in these regions may obscure the residues that determine these properties.
Comparison of the rates of synonymous and nonsynonymous substitutions in the V1 and V2 regions of env produced d_\text{s}/d_\text{s} ratios of 0.75, 0.73 and 0.97 for p4, p5 and p6 respectively. Similar values for the V1 and V2 regions were found in a previous study examining the relationship of HIV-1 infection between maternal and infant strains (Lamers et al., 1993). These ratios do not suggest strong positive or negative selection for sequence change acting on this region overall, being similar to previous estimates for the whole env gene (Wolfs et al., 1990; Li et al., 1988). The ratios are higher than previous estimates of around 0.4 in these study subjects (Hughes et al., 1997) and others (Gojobori et al., 1990b; Myers et al., 1992b; Li et al., 1988b; Kasper et al., 1995b) for p17\_gag, a region characterised by a high frequency of silent substitutions and conservative amino acid replacements. Ratios substantially greater than 1.0 imply positive selection, for example sequence comparisons of the V3 region has produced ratios of 1.5 to 2.9 (Lukashov et al., 1995; Simmonds et al., 1990). However, it is difficult to base conclusions for V1 and V2 on a single figure as the region probably contains a combination of functionally critical amino acids (such as the cysteine residues), those where variation is neutral in effect and those which may be subject to positive selection.

Immune recognition may play a major role in positive selection for antigenic variants of HIV-1, not only for cytotoxic T cell recognition (Phillips et al., 1991), but also from neutralising antibody recognising antigenic determinants in gp120. It has previously been suggested that the frequent insertions and deletions within V4 and V5 and the variability in the position and number of glycosylation sites may be mechanisms by which peptide epitopes are shielded from an evolving
immune response (Simmonds et al., 1990b). This hypothesis appears even more likely for V1 and V2, given the existence of neutralising epitopes in this region (Warrier et al., 1994; Ho et al., 1991; Yoshiyama et al., 1994; Moore et al., 1993; Gorny et al., 1994; Sullivan et al., 1993; Fung et al., 1992; van Tijn et al., 1989).

4.4.3 DOES THE DIVERSITY OF V1 AND V2 SEQUENCE VARIANTS CORRELATE WITH DISEASE PROGRESSION?

A method to compare populations of HIV variants by high resolution gel electrophoresis of DNA sequences amplified across the V4 and V5 hypervariable regions has previously been described (Simmonds et al., 1990b). Length profiles obtained in this way correlated well with the length obtained by sequencing of single molecules isolated at limiting dilution. In the current study, I applied the method of LPA to the V1 and V2 domains on variants amplified from non-lymphoid and lymphoid tissues from the three study subjects. Different length variants amplified from single molecules were equally represented in the analysis of length polymorphisms (Fig 20).

LPA provides only a partial description of the variability within a sample. Whereas each length variant represent a different amino acid sequence, often quite diverse sequences may have the same overall length, as was the case for p4 in V2 (Fig 21B). However, because such a large population can be amplified, less frequent variants visible as minor bands on the gel were frequently detected using LPA that were not detected by the amplification of single molecules, and in this
respect the method therefore provides a more complete analysis of diversity than sequencing based approaches (Table 9; Fig 20). Furthermore, LPA allows the rapid comparison of a large number of samples, providing a method to investigate the relationship between disease progression and tissue origin with population diversity on a greater range of samples than previously attempted. Although specific amino acid changes can not be detected by LPA, related techniques such as the heteroduplex mobility assays (Delwart et al., 1993) that have been used to investigate in vivo variability do not differentiate between the majority of substitutions that are silent from those that change amino acid sequences. Variability in these assays can therefore also not be precisely equated to changes that influence virus phenotype.

Previous studies have observed an association between the isolation of SI variants with an increased rate in disease progression (Tersmette et al., 1988; Richman et al., 1994; Connor et al., 1994). However, to date, sequence comparisons of variants in vivo have found little evidence for the existence of high frequencies of variants with an predicted SI phenotype based upon sequence comparisons in V3 (Lukashov et al., 1995; Wolinsky et al., 1996), including those of the study subjects described here (Donaldson et al., 1994a). Wolinsky et al., found no specific amino acid motifs associated with the appearance of variants with an SI phenotype in the V3 region from study subjects with evidence of both rapid T cell loss and stable T cell counts (Wolinsky et al., 1996). Therefore, although in vitro studies have suggested an increase in viral virulence may predetermine the rate of disease progression there seems to be little in vivo
evidence to corroborate this hypothesis.

The "antigenic diversity threshold" model was proposed to explain viral diversity and disease progression (see introduction, section 1.5.3; Nowak et al., 1991a). This theory predicts that during disease progression viral diversity is driven by an active immune response, resulting in escape mutants. This generation of antigenic diversity is thought to eventually overwhelm the immune system creating an anomaly whereby the active host immune response is intrinsically involved in its destruction. Wolinsky et al., have shown relatively homogenous populations to be present, over a period of approximately three years, in individuals with rapidly declining T cell counts (associated with rapid progression to disease) and comparably higher diversity in patients with stable T cell counts and moderately declining T cell counts (associated with a slower progression to disease; Wolinsky et al., 1996). A number of other studies have corroborated this finding, showing a decline in population diversity upon disease progression in regions of env that include V1 and V2 (McDonald et al., 1997; Lukashov et al., 1995; Delwart et al., 1994; Ganeshan et al., 1997). In this study, however I have shown that V1 and V2 viral variants were equally diverse in pre-symptomatic individuals and those dying in AIDS. I was therefore unable to experimentally confirm the findings of these previous studies for a decline in population diversity upon disease progression.

One of the difficulties of interpreting my own and published sequence comparisons is distinguishing between actively replicating HIV populations and those that may have infected cells latently or non-productively in the past. Evidence for variable persistence of non-expressing, "older" populations of HIV

213
have been obtained by comparisons of PBMC sequences in sequential samples from acutely infected individuals (where persistence may be extremely long; Simmonds et al., 1991), or by monitoring the appearance of resistant populations following antiviral treatment (where one third of the PBMC populations remained wild type 6 months after the onset of treatment; Wei et al., 1995). In these two studies, rapid and complete replacement of populations was observed only in the plasma virus population, which was generally less diverse than those in PBMCs for this reason. A recent study by Wong et al., found evidence for extremely slow turnover of variants infecting brain tissues from sequence comparisons of the pol gene (see chapter 3: section 3.6.4; Wong et al., 1997). Thus, particularly for the brain and other tissues where turnover of HIV populations may be slow, a clearer indication of the relationship between diversity and disease progression may be obtained by comparisons of the specifically transcriptionally active variants within the populations.

In summary, this study has shown no correlating factors between V1 and V2 with either an increase or decrease in diversity from pre-symptomatic and symptomatic individuals. No specific amino acid motifs were detected in either region which correlated with viral phenotype and no distinction could be made, from examining primary sequences, between variants found in different tissues. Therefore, there were no parameters in V1 or V2 that correlated with biological properties of the virus. If, as previously proposed, these two regions act in concert with other regions of gp120 to influence biological properties of HIV-1, this process remains undeterminable by analysis of their primary sequences.
CHAPTER 5: GENERAL DISCUSSION.
Since AIDS was first recognized in the early 1980s, the natural history of HIV-1 infection and clinical manifestations during progression of this disease have been elucidated. It is now recognized that HIV-1 infection induces a chronic and progressive disease process with a broad spectrum of clinical manifestations from acute primary infection to AIDS where a number of life threatening opportunistic infections and malignancies are observed. The course of disease is marked by increasing levels of viral replication, emergence of more virulent strains in vitro in approximately 50% of infected individuals, and progressive destruction of the immune system, primarily due to the resulting dysfunction and depletion of CD4 cells upon HIV-1 infection. Clinical AIDS is recognized as a combination of cytopathic infection, due to the direct action of HIV-1, and secondary or opportunistic infections primarily due to the suppression of the immune system. In the CNS, evidence of the cytopathic effect of HIV-1 occurs with increasing frequency and severity as immune defences are depleted (AIDS dementia complex/encephalitis), as do both minor (herpes zoster) and major (progressive multifocal leukoencephalopathy, toxoplasmosis and tuberculosis) opportunistic infections. This dichotomy of direct and indirect clinical manifestations is also observed in other tissues such as the gastrointestinal system, liver and heart during disease progression.

In order to investigate whether the spread of HIV-1 infection from lymphoid to non-lymphoid tissues occurs prior to immunosuppression (as an early event) or after an individual has progressed to full blown AIDS (a late event) I
have analysed the nucleic acid sequences of the p17gag region from a number of lymphoid and non-lymphoid tissues obtained from individuals dying in AIDS. Also in order to determine whether the V1 or V2 hypervariable regions determine infection of such tissues I have analysed both primary amino acid sequences and nucleic acid sequences from tissue samples from the same patients. During disease progression an increase in infected tissues and clinical manifestations is evident. To investigate whether this is related to the degree of diversity of HIV-1 I have analysed brain and lymph node tissue from a wide range of symptomatic and pre-symptomatic patients.

Phylogenetic analysis of the p17gag region revealed polyphyletic lineages of lymphoid and non-lymphoid tissues with obvious mixing between all tissues examined in three of the four patients analysed (p4, 5 and 6). The observation of polyphyletic lineages in brain provides no evidence for a specifically neurotropic variant of HIV, adapted for infection of brain tissue, where the expectation would be for variants recovered from brain to be monophyletic.

One of the surprising findings in the course of this study was the high diversity of the viral population in the brain. Mean synonymous pairwise distances were calculated using the p17gag region to estimate the average time of divergence of variants within each tissue. The average time of divergence between sequences from brain from patient 4, 5 and 6 (4.1 to 6.5 years) was found to be significantly higher than that estimated for lymphoid variants (2.65 to 6.5 years) suggesting infection of brain occurred relatively early in infection.

The restricted diversity observed in lymphoid tissue suggests that complete
population replacements may have occurred due to a more rapidly diverging population. This hypothesis is compatible with recent findings which have shown a rapid turnover of virus and CD4 lymphocytes in the peripheral circulation throughout infection (Ho et al., 1995; Wei et al., 1995). However, diversification of variants within brain tissue may be facilitated by the lack of cellular mobility, while the absence of positive selective pressure for change may continue to restrict sequence variability (see section 3.6.1).

Within the V1 and V2 regions the lack of tissue specific sequences was reflected in the phylogenetic analysis which revealed a number of mixed lineages containing variants from both lymphoid and non-lymphoid tissues. Further analysis revealed no association between charge, length or number of glycosylation sites with tissue origin, consistent with the findings of Wang et al., (Wang et al., 1995). Therefore, there would appear to very little direct evidence from examination of primary sequences alone to substantiate that the V1/V2 regions influence viral tropism. Similarly, the claims that infection of brain tissue may require specific neurotropic variants would appear to be unfounded from examination of these regions. Amino acid differences in V1 and V2 between brain and lymphoid tissues noted in two of the three patients (p4 and 5) may be due to a lack of divergence as a result of the absence of selective pressure from the immune system or may simply reflect a requirement for infection of macrophage-like cells. In vitro studies have shown that these two regions may influence infection of macrophages to some extent (see section 4.1). However, this property may be influenced by interactions with other regions of gp120, and hence may be invisible from the examination of
primary sequences.

Analysis of the V3 region has suggested that this region may also be an important determinant of HIV-1 tropism (see section 1.4.3). Takeuchi et al., reported the isolation of an HIV-1 variant that infected CD4+ brain cells (HIV\textsubscript{GUN}), and showed that a single point mutation of the highly conserved proline at the tip of the V3 loop (GPGR) conferred an ability on this isolate to replicate in fibroblast like cells derived from human brain (BT cells; Takeuchi et al., 1991). A subsequent study by the same research group isolated several new variants able to infect brain derived cells with a mutation from proline to serine, threonine or alanine in the conserved GPGR sequence at the tip of the V3 loop, suggesting amino acid sequences in this region may be important for infection of brain (Shimizu et al., 1994). However, in an analysis of \textit{in vivo} V3 sequences isolated from lymphoid and non-lymphoid tissues a proline residue was observed in the tip of the V3 loop, which was conserved in all sequences, although in three of the patients (one asymptomatic and two symptomatic) GPGS was detected in a number of sequences (Donaldson et al., 1994a). In the same study the major variants found in brain tissue were also detected as minor variants in other tissues with relatively trivial differences detected between variants from different tissues, consisting of only one or two amino acids. Therefore, similar to my findings from analysis of the V1 and V2 sequences, there appears to be no signature sequence in the V3 region to indicate the existence of a specific neurotropic variant.

Analysis of primary sequences is a useful tool when constructing evolutionary relationships and provides important information regarding the
divergence of viral variants enabling us to discern when various tissues became infected during disease progression. However, analysis of primary sequences provides little information about the biological properties of a virus as these sequences are treated as distinct variables and cannot be analysed with respect to their secondary and tertiary interactions with other regions of the viral genome. Although such interactions may be speculated upon, no definite conclusions can be made from the examination of primary sequences alone. In the future, it is likely that more substantial evidence will require the functional characterization of HIV variants cloned directly from different tissues (see below).

The finding of high diversity within brain is in accordance with previous studies that suggest HIV can be detected in the brain at all stages of disease and in individuals without GCE (Davis et al., 1992). In contrast, a number of other investigations have suggested a lack of productive infection of non-lymphoid tissues prior to immunosuppression where significant levels of provirus in non-lymphoid tissues from asymptomatic patients were not detected (Donaldson et al., 1994b; Bell et al., 1993). In a number of patients low levels of provirus were detected, although these levels were consistent with that expected from contamination by residual infected blood within tissues. This contrasts the detection of substantial levels of provirus in tissues from patients with AIDS, which were several orders of magnitude greater than that detected in asymptomatic patients. Furthermore, pathological examination of brain from asymptomatic patients failed to reveal any evidence for HIV encephalitis and p24 antigen was not detected by immunocytochemistry (Bell et al., 1993). In a subsequent larger study, a clear
relationship between HIV encephalitis, detection of p24 antigen by immunocytochemistry and proviral load in brain from AIDS patients was detected (Bell et al., 1996). More recently in situ PCR has been developed for detection of proviral sequences in tissue, and the results confirm previous findings of an absence of detectable infected cells in asymptomatic patients, with frequent detection of infected cells in AIDS patients with GCE (Bell et al., 1996).

Therefore, infection of brain tissue may not be productive during the asymptomatic stage of infection, it may be present at low levels undetectable by present techniques or be so dispersed in the brain that infection is not apparent from the single samples normally used for PCR, or histological examination.

The mode of entry of HIV-1 into brain is not well understood. It has been postulated that HIV-1 infected macrophages from the peripheral blood may be able to cross the blood-brain barrier bringing with them the virus ("Trojan horse" mechanism). This hypothesis is supported by the common origin and phenotype of macrophages and microglia. The restricted sequence diversity in V3 suggests that infection of brain simply may be a consequence of a requirement for macrophage-tropism and does not represent a specifically adapted neurotropic variant. Indeed, the existence of specific neurotropic variants would suggest that sequences of HIV from the brain should be monophyletic, whereas this was clearly not the case in most of the study subjects described here. Early seeding of brain (and other non-lymphoid tissues) may occur as the result of an inadequate immune response during primary infection with HIV-1. However, as an efficient immune response is mounted, viral replication within tissues may be contained at a very low level.
during the asymptomatic stage. Subsequently, as disease progresses viral replication within tissues may no longer be limited and result in the pathological and clinical manifestations recognized in AIDS patients. In this respect, HIV-1 may be similar to the ubiquitous JC virus which causes a subacute demyelinating disease, progressive multifocal leukoencephalopathy (PML), infecting primarily oligodendrocytes and astrocytes. Seroepidemiological studies have shown approximately 70% of the population possess antibodies to JC virus (Padgett et al., 1973), although infection remains undetectable in normal brain by modern techniques. PML occurs almost exclusively in immunosuppressed individuals, most often associated with lymphoproliferative disorders such as leukaemias and lymphomas, although in recent years PML has become increasingly common amongst AIDS patients, with the estimated incidence to be in the order of 3.8% (Berger et al., 1987). As clinical manifestations of JC virus infection occur only in the absence of an effective host immune response, as appears to be the case for infection with HIV, reactivation of the latent virus may be the cause.

In order to understand the biological properties of a virus, functional studies must be carried out to examine these properties in as close to the natural state as possible. These generally involve in vitro culture of the virus in lymphocytes followed by characterization of the variants present. However, this technique can be misleading. It has been established for some time that culturing of HIV-1 selects for the outgrowth of more virulent variants in vitro which leads to the misrepresentation of viral populations. Also, during culturing of variants from tissues such as brain, bowel and lung there is a high probability of contamination.
from residual lymphocytes which may be preferentially isolated upon co-culture with PBMCs, particularly if there were marked differences in tropism in tissue adapted variants. Even amongst variants present in PBMCs, there is evidence for strong selection for SI variants within the first week of \textit{in vitro} culture even where the majority of the PBMC population has a predicted NSI macrophage tropic phenotype (P. Strappe, personal communication). In this study, all of the SI variants replicated in MT-2 cells and in each case the variants selected for had a higher charge (+5) than PBMC populations (+3).

We have recently attempted to address the problem of \textit{in vitro} selection through a collaboration with Matthias Dittmar and colleagues (Dittmar \textit{et al.}, 1997), in which variants of HIV infecting different tissues were cloned and expressed without prior \textit{in vitro} culture. Long range PCR was used to construct a full-length provirus sequence from a primary isolate of p4 (one of the study subjects in this thesis) obtained three months before death, into which amplified \textit{env} sequences from lymphoid and non-lymphoid tissues from the same study subject were inserted. The resultant recombinant viruses were characterized for biological properties such as tropism, cytopathology and co-receptor usage. This method eliminates \textit{in vitro} selection of unrepresentative populations, since these replication competent viruses have been constructed from genomic DNA amplified directly from tissue samples. Remarkably, initial biological characterization of the expressed viruses from brain, lung and lymph node revealed that irrespective of tissue origin, they could only productively infect peripheral blood mononuclear T cells (Dittmar \textit{et al.}, 1997). In contrast, recombinant viruses comprising the p4
background sequence with inserted env sequences from HIV$_{gun}$ and HIV$_{SF162}$ (macrophage tropic variants) were able to infect primary macrophages, ruling out the possibility that macrophage tropism was determined by parts of the genome of this clone outwith the env gene.

Furthermore, infection of PBMC with these recombinant viruses showed similar replication kinetics regardless of tissue origin. These findings now cast a shadow of doubt on the hypothesis that the only requirement for infection of brain is the ability to infect macrophages, and it is possible that there may be other factors which can influence viral tropism for tissues of non-lymphoid origin such as brain and lung. Although microglia are considered to be macrophage derived, a recent in vitro study has suggested that certain isolates may replicate preferentially in microglia (Strizki et al., 1996). Extensive passaging of such an isolate revealed amino acid changes in the V3 loop shown to be associated with isolates from patients with HIV dementia.

Matthias Dittmar and colleagues are currently examining the possibility that co-receptor usage may influence infection of tissues such as brain. They intend to analyse the chemokine receptor usage of variants from different tissues, such as brain and lymph node, for entry into different cell types. In order to do this a number of proviruses, generated by long range PCR, from brain and lymph node tissue obtained from the same patient will be directly cloned and used to infect various different in vitro cell systems such as primary macrophages, primary dendritic cells and CD4+ lymphocytes.

In the future, functional and genetic characterization studies may provide
more insight with regard to the disease manifestations associated with HIV infection. Understanding the pathogenesis of HIV infection will involve the reconciliation of both clinical and pathological observations with virological studies. Infection of non-lymphoid tissues, such as the brain, may be a consequence of alterations in the character of the virus (tropism), discussed above, or a result of the ensuing immunosuppression characteristic of HIV infection. The fact that pathological abnormalities in non-lymphoid tissues, such as HIV encephalitis occur in the setting of immunosuppression and more active viral replication suggests that a decline in host defences against HIV may be an important factor in allowing infection of non-lymphoid tissues. Therefore, there is strong evidence to suggest that the host response may be instrumental in the containment of viral spread during the asymptomatic stage of HIV infection.


CORNELISSEN, M., HOGERVORST, E., ZORGDRAGER, F., HARTMAN, S. &


DIETRICH, U., GREZ, M., VONBRIESEN, H., PANHANS, B.,
GEISSENDORFER, M., KUHNEL, H., MANIAR, J., MAHAMBRE, G.,
HIV-1 strains from India are highly divergent from prototypic African and
United-States/European strains, but are linked to a South African isolate - short
communication. AIDS 7, 23-27.


DITTMAR, M.T., SIMMONS, G., DONALDSON, Y.K., SIMMONDS, P.,
of human immunodeficiency virus type 1 (HIV-1) clones derived from different
submitted.

DITZEL, H.J., BINLEY, J.M., MOORE, J.P., SODROSKI, J., SULLIVAN, N.,
SAWYER, L.S.W., HENDRY, R.M., YANG, W.P., BARBAS, C.F. & BURTON,
D.R. (1995). Neutralizing recombinant human antibodies to a conformational V2-
and CD4-binding site-sensitive epitope of HIV-1 gp120 isolated by using an

infection of CD4+ human lung fibroblasts. AIDS 6, 232-233.

DOMINGO, E., DAVILA, M. & ORTIN, J. (1980). Nucleotide sequence
heterogeneity of the RNA from a natural population of foot and mouth disease

DONALDSON, Y.K., BELL, J.E., HOLMES, E.C., HUGHES, E.S., BROWN,
of human immunodeficiency virus type 1 showing restricted sequence variability in
the V3 loop. J.Virol. 68, 5991-6005.

DONALDSON, Y.K., BELL, J.E., IRONSIDE, J.W., BRETTLE, R.P.,
HIV outside the lymphoid system with onset of AIDS. Lancet 343, 382-385.

DORFMAN, T., MAMMANO, F., HASELTINE, W.A. & GOTTLINGER, H.G.
(1994). Role of the matrix protein in the virion association of the human
immunodeficiency virus type 1 envelope glycoprotein. J.Virol. 68, 1689-1696.

DORNER, A.J. & COFFIN, J.M. (1986). Determinants for receptor interaction and
cell killing on the avian retrovirus glycoprotein gp85. Cell. 45, 365-373.


antigens associated with human T-cell leukemia virus patients with AIDS. Science 220, 859-862.


246


GORNİ, M.K., CONLEY, A.J., KARWOWSKA, S., BUCHBINDER, A., XU,

247


248


KANG, C.Y., NARA, P., CHAMAT, S., CARALLI, V., RYSKAMP, T.,
non-V3-specific neutralizing antibodies that interfere with gp120/CD4 binding in
6171-6175.

KANKI, P.J., KURTH, R., BECKER, W., DREISMAN, G., MOLANE, M.F. &
essex, M. (1985). Antibodies to simian T-lymphotropic retrovirus type III in
African green monkeys and recognition of STLV-III viral proteins by AIDS and
related sera. Lancet i, 1330-1332.

Human immunodeficiency virus type 2 vpx protein augments viral infectivity.
Virology 184, 197-209.

KAPPLER, J., KOTZIN, B., HERRON, L., GELFAND, E., BIGLER, R.,
BOYLSTON, A., CARREL, S., POSNETT, D., CHOI, Y. & MARRACK, P.
Science. 244, 811-813.

RNA binding by the tat and rev proteins of HIV-1. Biochimie 73, 9-16.

KASPER, P., KAISER, R., OLDENBURG, J., BRACKMANN, H.H., MATZ, B. &
schneweis, K.E. (1994). Parallel evolution in the V3 region of HIV type 1 after
infection of hemophiliacs from a homogeneous source. AIDS Res.Hum.Retroviruses
10, 1669-1678.

KASPER, P., SIMMONDS, P., SCHNEWEIS, K.E., KAISER, R., MATZ, B.,
diversification of the HIV type 1 gag p17 gene in patients infected from a common

Antibody dependent cellular cytotoxicity (ADCC)- mediated destruction of human
immunodeficiency virus (HIV) coated CD4+ T lymphocytes by acquired immune


substitution through comparative studies of nucleotide sequences. J. Mol. Evol. 16,
111-120.


258


LOUWAGIE, J., MCCUTCHAN, F.E., PEETERS, M., BRENAN, T.P.,


MACREADIE, I.G., CASTELLI, L.A., HEWISH, D.R., KIREPATRIDE, A.,


NOWAK, M.A., ANDERSON, R.M., MCLEAN, A.R., WOLFS, T.F.W.,


273


274


HIV-1 RNA. *Science.* 247, 1531


An inducible mammalian amber suppressor: propagation of a poxvirus mutant. 


SMITH, D.D., BYRN, R.A., MARSTERS, S.A., GREGORY, T., GROOPMAN,


282


modification of cell tropism by a single point mutation at the neutralization epitope in the env gene. J.Virol. 65, 1710-1718.


WANG, W.K., ESSEX, M. & LEE, T.H. (1996). Single amino acid substitution in constant region 1 or 4 of gp120 causes the phenotype of a human immunodeficiency virus type 1 variant with mutations in hypervariable regions 1 and 2 to revert. J.Virol. 70, 607-611.


APPENDICES
Appendix I: Specified secondary infectious diseases listed in the CDC surveillance definition for AIDS (CDC, 1986).

This group includes patients with symptomatic or invasive disease due to one of twelve specified secondary infections listed below.

- *Pneumocystis carinii* pneumonia
- Chronic cryptosporidiosis
- Toxoplasmosis
- Extraintestinal strongyloidiasis
- Isosporiasis
- Cryptococcosis
- Histoplasmosis
- Mycobacterial infection (*Mycobacterium avium*)
- Chronic mucocutaneous
- Disseminated herpes simplex virus infection
- Progressive multifocal leukoencephalopathy
Appendix II: Other specified secondary infectious diseases (CDC., 1986).

This group includes patients with one of six other specified secondary infectious diseases listed below.

. Oral hairy leukoplakia
. Multidermatomal herpes zoster
. Recurrent salmonella bactaeremia
. Nocardiosis
. Tuberculosis
. Oral candidiasis (thrush)
Appendix III: Secondary cancers (CDC., 1986).

This group includes patients with one or more of the specified cancers listed below.

. Karposi’s sarcoma
. non Hodgkin’s lymphoma
. Primary lymphoma of the brain
. Cervical cancer
Appendix IV: Nucleotide sequences obtained for p79.

P79BR-3A  GGG AGC TAG AAC GAT TCG CAG TTA ACC CTG GCC TAT  
TAG AAA CAT CAG AAG GCT GTA GAC AAA TAT TGG AAC AGC TAC  
AAC CAT CCC TTC AGA CAG GATCAA AAG AAC TTA GAT CAT TAT  
ATA ATG CAA TAG CAA CTC TCT ATT GTG TGC ATC AAA GGA TAG  
AGC AAA ACA AAA ACC AGG AAA AAG CAC AGC AAG ACG AAG  
CAG CAG CTT GCA CAG GAA ACA GCA GTC AGG CCA GCC AAA ATT  
ACC CTA TAG TGC AAA ACA GCC AGG ???? ??? ??? ??? ??? ???  

P79BR-5A  GGG AGC TAG AAC GAT TCG CAG TTA ACC CTG GCC TAT  
TAG AAA CAT CAG AAG GCT GTA GAC AAA TAT TGG AAC AGC TAC  
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAC TTA GAT CAT TAT  
ATA ATG CAG TAG CAA CTC TCT ATT GTG TGC ATC AAA GGA TAG  
AGC AAA ACA AAA ACC AGG AAA AAG CAC AGC AAG ACG AAG  
CAG CAG CTT GCA CAG GAA ACA GCA GTC AGG CCA GCC AAA ATT  
ACC CTA TAG TGC AAA ACA GCC AGG ???? ??? ??? ??? ??? ???  

P79BR-7A  GGG AGC TAG AAC GAT TCG CAG TTA ACC CTG GCC TAT  
TAG AGA CAT CAG AAG GCT GTA GAC AAA TAT TGG AAC AGC TAC  
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAC TTA GAT CAT TAT  
ATA ATG CAG TAG CAA CTC TCT ATT GTG TGC ATC AAA GGA TAG  
AGC AAA ACA AAA ATC AGG AAA AAG CAC AGC AAG CAC AGG  
CAG CAG CTT GCA CAG GAA ACA GCA GTC AGG CCA GCC AAA ATT  
ACC CTA TAG TGC AAA ACA GCC AGG ???? ??? ??? ??? ??? ???  

P79BR-8A  GGG AGC TAG AAC GAT TCG CAG TTA ACC CTG GCC TAT  
TAG AAA CAT CAG AAG GCT GTA GAC AAA TAT TGG AAC AGC TAC  
AAC CAT CCC TTC AGA CAG GAT CAA AAG AAC TTA GAT CAT TAT  
ATA ATG CAG TAG CAA CTC TCT ATT GTG TGC ATC AAA GGA TAG  
AGC AAA ACA AAA ACC AGG AAA AAG CAC AGC AAG CAC AGG  
CAG CAG CTT GCA CAG GAA ACA GCA GTC AGG CCA GCC AAA ATT  
ACC CTA TAG TGC AAA ACA GCC AGG ???? ??? ??? ??? ??? ???  

P79BR-9A  GGG AGC TAG AAC GAT TCG CAG TTA ACC CTG GCC TAT  
TAG AAA CAT CAG AAG GCT GTA GAC AAA TAT TGG AAC AGC TAC  
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAC TTA GAT CAT TAT  
ATA ATG CAG TAG CAA CTC TCT ATT GTG TGC ATC AAA GGA TAG  
AGC AAA ACA AAA ATC AGG AAA AAG CAC AGC AAG CAC AGG  
CAG CAG CTT GCA CAG GAA ACA GCA GTC AGG CCA GCC AAA ATT  
ACC CTA TAG TGC AAA ACA GCC AGG ???? ??? ??? ??? ??? ???  

296
TAG AGA CAT CAG AAG GCT GTA GAC AAA TAT TGG AAC AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAC TTA GAT CAT TAT
TTA ATA CAG TAG CAA CCC TCT ATT GTG TGC ATC AAA AGA TAG
ATG TAA AAG ACA CCA ATG AAG CTT TAG AGA AGA TAA AGG AAG
AGC CAG CAG CTA GCA GCA AGA ACA GCA GTC AGG CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TTC AGG GGC AAA ??? TGG ??? ???

P79BR-53A GGG AGC TAG AAC GAT TCG CAG TTA ACC CTG GCC TAT
TAG AAA CAT CAG AAG GCT GTA GAC AAA TAT TGG AAC AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAC TTA GAT CAT TAT
ATA ATG CAG TAG CAA CTC TCT ATT GTG TGC ATC AAA GGA TAG
ATG TAA AAG ACA CCC AGG AAG CTT TAG AGA AGA TAA AGG AAG
AGC AAA ACA AAA ATC AGG AAA AAG CAC AGC AAG CAC AGC AAG
CAG CAG CAG GCA CAG GAA ACA GCA GTC AGG CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TTC AGG GGC AAA TGG TAC ATC AGG
CCA TA? ??? ??? ??? ??? ??? ??? ???

P79BR-54A GGG AGC TAG AAC GAT TCG CAG TTA ACC CTG GCC TAT
TAG AAA CAT CAG AAG GCT GTA GAC AAA TAT TGG AAC AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAC TTA GAT CAT TAT
ATA ATG CAG TAG CAA CTC TCT ATT GTG TGC ATC AAA GGA TAG
ATG TAA AAG ACA CCC AGG AAG CTT TAG AGA AGA TAA AGG AAG
AGC AAA ACA AAA ATC AGG AAA AAG CAC AGC AAG CAC AGC AAG
CAG CAG CAG GCA CAG GAA ACA GCA GTC AGG CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TTC AGG GGC AAA TGG TAC ATC AGG
CCA TA? ??? ??? ??? ??? ??? ??? ???

P79BR-60A GGG AGC TAG AAC GAT TTG CAG TTA ACC CTG GCC TAT
TAG AAA CAT CAG AAG GCT GTA GAC AAA TAT TGG AAC AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAC TTA GAT CAT TAT
ATA ATG CAG TAG CAA CTC TCT ATT GTG TGC ATC AAA GGA TAG
ATG TAA AAG ACA CCC AGG AAG CTT TAG AGA AGA TAA AGG AAG
AGC AAA ACA AAA ATC AGG AAA AAG CAC AGC AAG CAC AGC AAG
CAG CAG CAG GCA CAG GAA ACA GCA GTC AGG CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TTC AGG GGC AAA TGG TAC ATC AGG
CCA TA? ??? ??? ??? ??? ??? ??? ???

P79LN-4B GGG AAC TAG AAC GAT TCG CAG TCA ACC CTG GCC TAT
TAG AAA CAT CAG AAG GCT GTA GAC AAA TAC TGG AAC AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAC TTA GAT CAT TAT
ATA ATG CAG TAG CAA CCC TCT TTG GTG TGC ATC AAA GGA TAG
ATG TAA AAG ACACCC AGG AAG CTT TAG AGA AGA TAA AGG AAG
AGC AAG ACA AAA GTA AGA AAA AAG CAC AGC AAG CAC AGC AAG
CAG CAG CAG GCA CAG GAA ACA GCA GTC AGG CCA GCC AAA ATT
ACC CTA TAG TGC AGA ACA TTC AGG GGC AAA TGG TAC ATC AGG
CAA TAT AAA CCT A? ??? ??? ??? ??? ???

P79LN-8B ??? ??? ??? ??? ??? ??? TCG CAG TTA ACC CTG GCC TAT
TAG AAA CAT CAG AAG GCT ATA GAC AAA TAC TGG AAC AGC TAC
GGC CAT CCC TTC AGA CAG GAT CAG AAG AAC TTA GAT CAT TAT
ATA ATG CAG TAG CAA CTC TCT ATT GTG TGC ATC AAA GGA TAG
ATG TAA AAG ACA CCC AGG AAG CTT TAG AGA AGA TAA AGG AAG
AGC AAA GCA AAA GTC AGA AAA AAG CAC AGC AAG CAC AGC TAG
TAG CAG CTA GCA GAG AAC ACA GCG GTC AGG CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TCC AGG AGC TAA TGC TAT GCC AAG
???? ??? ??? ??? ??? ??? ??? ??? ???
P79LN-13B GGG AGC TAG AAC GAT TCG CAG TTA ACC CTG GCC TAT
TAG AAA CAT CAG AAG GCT GTG GAC AAA TAT TGG AAC AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAC ATT TTA GAT CAT TAT
ATA ATG CAT TAG CAA CTC TCT ATT GTG TGG ATC AAA GGA TAG
ATG TAA AAG ACA CCC AGG AAG CTT TAG AGA AGA TAA AGG AAG
AGC AAA ACA AAA ATC AGA AAA AAG CAC AGC AAG CAC AGC AAG
CAG CAG CTC GCA CAG GAA ACA GCA GTC AGG CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TCC AGG AGG AAA TGG TAC ATC AGG
CCA TAT A?? ??? ??? ??? ??? ??? ???
P79LN-15B GGG AGC TAG AAC GAT TTG CAG TTA ACC CTG GCC TCT
TAG AAA CAT TAG AAG GCT GTG GAC AAA TAT TGG AAC AGC TAC
AGC CAT CCC TTC AGA CAG GAT CAG AAG AAC TTA GAT CAT TAT
ATA ATG CAT TAG CAA CTC TCT ATT GTG TGG ATC AAA GGA TAG
ATG TAA AAG ACA CCC AGG AAG CTT TAG AGA AGA TAA AGG AAG
AGC AAA GCA AAA GTC AGA AAA AAG CAC AAC AAG CAC AGC AAG
CAG CAG CTC GCA CAG GAA ACA GCA GCG GTC AGG CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TCC AGG AGG AAA TGG TAC ATC AGG
CCA TAT A?? ??? ??? ??? ??? ??? ???
P79LN-20B GGG AGC TAG AAC GAT TCG CAG TCA ACC CTG GCC TAT
TAG AGA CAT CAG AAG GCT GTG GAC AAA TAT TGG AAC AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAC TTA GAT CAT TAT
ATA ATG CAT TAG CAA CTC TCT ATT GTG TGG ATC AAA GGA TAG
ATG TAA AAG ACA CCC AGG AAG CTT TAG AGA AGA TAA AGG AAG
AGC AAA GCA AAA GTC AGA AAA AAG CAC AAC AAG CAC AGC AAG
CAG CAG CTC GCA CAG GAA ACA GCA GCG GTC AGG CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TCC AGG AGG AAA TGG TAC ATC AGG
CCA TAT A?? ??? ??? ??? ??? ??? ???
P79LN-50B GGG AGT TAG AAC GAT TCG CAG TTA ACC CTA GCC TAT
TAG AAA CAT CAG AAG GCT GTG GAC AAA TAT TGG AAC AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAC ATT TTA GAT CAT TAT
ATA ATG CAT TAG CAA CTC TCT ATT GTG TGG ATC AAA GGA TAG
AGG TAA AAG ACA CCC AGG AAG CTT TAG AGA AGA TAA AGG AAG
AGC AAA GCA AAA GTC AGA AAA AAG CAC AAC AAG CAC AGC AAG
CAG CAG CTC GCA CAG GAA ACA GCA GCG GTC AGG CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TCC AGG ??? ??? ??? ??? ??? ???
P79LN-56B GGG AGT TAG AAC GAT TCG CAG TTA ACC CTG GCC TAT
TAG AAA CAT CAG AAG GCT GTG GAC AAA TAT TGG AAC AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAC TTA GAT CAT TAT
ATA ATG CAT TAG CAA CTC TCT ATT GTG TGG ATC AAA GGA TAG
AGG TAA AAG ACA CCC AGG AAG CTT TAG AGA AGA TAA AGG AAG
AGC AAA GCA AAA GTC AGA AAA AAG CAC AGC AAG CAC AGC AAG
CAG CAG CTC GCA CAG GAA ACA GCA GCG GTC AGG CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TCC AGG ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ??? ???
ATA ATG CAG TAG CAA CTC TCT ATT GTG TGC ATC AAA GGA TAG
AGG TAA AAG ACA CCC AGG AAG CTT TAG AGA AGA TAA AGG AAG
AAC ACA GCA AAA GTC AGA AAA AAG CAC AGC AAG CAC AGC TAG
TAG CAG CTG ACA CAG GAA ACA GCA GTC AGG CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TCC AGG GGC AAA TGG TAC ATC ??
?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ??

P79LG-4C GGG AGC TAG AAC GAT TCG CAG TTA ACC CTG GCC TAT
TAA AAA CAT CAG AAG GCT GTA GAC AAA TAC TGG AAC AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAC TTA TAT CAT TAT
TTA ATA CAG TAG CAA CCC TCT ATT GTG TGC ATC AAA GGA TAG
ATG TAA AAG ACA CCA ATG AAG CTT TAG AGA AGA TAA AGA AAA
AGC AAA ACA AAA GTC AGA AAA AAG CAC AGC AAG CAC AGC AGG
CAG CAG CTG GCA CAG GAA ACA GCA GTC AGA CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TCC AGG GGC AAA TGG TAC A?? ??
?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ??

P79LG-9C GAG AGC TAG AAC GAT TCG CAG TTA ACC CTG GCC TAT
TAG AAG CAT CAG AAG GCT GTA GAC AAA TAT TGG AAC AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAC TTA TAT CAT TAT
ATA ATA CAG TAG CAA CCC TCT ATT GTG TGC ATC AAA GGA TAG
AGG TAA AAG ACA CCA ATG AAG CTT TAG AGA AGA TAA AGG AAA
AGC AAG ACA AAG GTC AGA AAA AAG CAC AGC AAG CAC AGC AGG
CAG CAG CTG GCA CAG GAA ACA GCA GTC AGA CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TCC AGG GGC AAA TGG TAC A?? ??
?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ??

P79LG-10C GGG AGC TAG AAC GAT TCG CAG TTA ACC CTG GCC TAT
TAG AAG CAT CAG AAG GCT GTA GAC AAA TAT TGG AAC AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAC TTA TAT CAT TAT
ATA ATA CAG TAG CAA CCC TCT ATT GTG TGC ATC AAA GGA TAG
ATG TAA AAG ACA CCA GGG AAG CTG TAC TGT TGC ATC AAA GGA TAG
AGC AAG ACA AAG GTC AGA AAA AAG CAC AGC AAG CAC AGC AGG
CAG CAG CTG GCA CAG GAA ACA GCA GTC AGA CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TCC AGG GGC AAA TGG TAC A?? ??
?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ??

P79LG-13C GGG AGC TAG AAC GAT TCG CAG TTA ACC CTG GCC TAT
TAG AAG CAT CAG AAG GCT GTA GAC AAA TAT TGG AAC AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAC TTA TAT CAT TAT
ATA ATA CAG TAG CAA CCC TCT ATT GTG TGC ATC AAA GGA TAG
ATG TAA AAG ACA CCA GGG AAG CTG TAc TGT TGC ATC AAA GGA TAG
AGC AAG ACA AAG GTC AGA AAA AAG CAC AGC AAG CAC AGC AGG
CAG CAG CTG GCA CAG GAA ACA GCA GTC AGA CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TCC AGG GGC AAA TGG TAC A?? ??
?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ??

P79LG-15C GGG AGC TAG AAC GAT TCG CAG TTA ACC CTG GCC TAT
TAG AAG CAT CAG AAG GCT GTA GAC AAA TAT TGG AAC AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAC TTA TAT CAT TAT
ATA ATA CAG TAG CAA CCC TCT ATT GTG TGC ATC AAA GGA TAG
ATG TAA AAG ACA CCA GGG AAG CTG TAC TGT TGC ATC AAA GGA TAG
AGC AAG ACA AAG GTC AGA AAA AAG CAC AGC AAG CAC AGC AGG
AGC AAA ACA AAG GTC AGA AAA AAG CAC AGC AAG CAC AGC AGG

302
CAG CAG CTG GCA CAG GAA ACA GCG GTC AGG CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TCC AGG GGC AAA TGG TAC ??? ???

P79LG-25C GAT AGC TAG AAG GAT TCG CAG TTA ACC CTG GCC TAT
TAG AAG CAT CAG AAG GCT GTA GAC AAA TAT TGG AAG AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAG TTA GAT CAT TAT
ATA ATG CAG TAG CAA CTC TTC ATT GTG TGA ATC AAA GAA TAG
ATG TAA AAG ACA CCA GGC AGG CTG TAG AGA AGA TAA AGG AAG
AGC AAA GCA AAA GTC AGA AAA AAG CAG AGC AAA CAC AGC AAG
CAG CAG CTG GCA CAG AAG GGC GTC AGG CCA GCC AAA ATT
ACC CCA TAG TGC AAA ACA TCC AGG GGC AAA TGG TA? ??? ???

P79LG-29C GAG AGC TAG AAC GAT TCG CAG TTA ACC CTA GCC TAT
TAG AGA CAT CAG AAG GCT GTA GAC AAA TAT TGG AAG AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAG TTA GAT CAT TAT
ATA ATG CAG TAG CAA CTC TTC ATT GTG TGA ATC AAA GAA TAG
ATG TAA AAG ACA CCA GGC AGG CTG TAG AGA AGA TAA AGG AAG
AGC AAA GCA AAA GTC AGA AAA AAG CAG AGC AAA CAC AGC AAG
CAG CAG CTG GCA CAG AAG GGC GTC AGG CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TCC AGG GGC AAA TGG TA? ??? ???

P79LG-33C GGG AGC TAG AAC GAT TCG CAG TTA ACC CTG GCC TAT
TAG AAG CAT CAG AAG GCT GTA GAC AAA TAT TGG AAG AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAG TTA GAT CAT TAT
ATA ATG CAG TAG CAA CTC TTC ATT GTG TGA ATC AAA GAA TAG
ATG TAA AAG ACA CCA GGC AGG CTG TAG AGA AGA TAA AGG AAG
AGC AAA GCA AAA GTC AGA AAA AAG CAG AGC AAA CAC AGC AAG
CAG CAG CTG GCA CAG AAG GGC GTC AGG CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TCC AGG GGC AAA TGG TA? ??? ???

P79LG-36C GGG AGC TAG AAC GAT TCG CAG TTA ACC CTG GCC TAT
TAG AAG CAT CAG AAG GCT GTA GAC AAA TAT TGG AAG AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAG TTA GAT CAT TAT
ATA ATG CAG TAG CAA CTC TTC ATT GTG TGA ATC AAA GAA TAG
ATG TAA AAG ACA CCA GGC AGG CTG TAG AGA AGA TAA AGG AAG
AGC AAA GCA AAA GTC AGA AAA AAG CAG AGC AAA CAC AGC AAG
CAG CAG CTG GCA CAG AAG GGC GTC AGG CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TCC AGG GGC AAA TGG TA? ??? ???

P79LG-38C GGG AGC TAG AAC GAT TCG CAG TTA ACC CTG GCC TAT
TAG AAG CAT CAG AAG GCT GTA GAC AAA TAT TGG AAG AGC TAC
AAC CAT CCC TTC AGA CAG GAT CAG AAG AAG TTA GAT CAT TAT
ATA ATG CAG TAG CAA CTC TTC ATT GTG TGA ATC AAA GAA TAG
ATG TAA AAG ACA CCA GGC AGG CTG TAG AGA AGA TAA AGG AAG
AGC AAA GCA AAA GTC AGA AAA AAG CAG AGC AAA CAC AGC AAG
CAG CAG CTG GCA CAG AAG GGC GTC AGG CCA GCC AAA ATT
ACC CTA TAG TGC AAA ACA TCC AGG GGC AAA TGG TA? ??? ???

303
## Appendix V

<table>
<thead>
<tr>
<th>Data Origin</th>
<th>Sequence</th>
<th>Phenotype</th>
<th>V1 Charge</th>
<th>V1 Length</th>
<th>V1 CHO Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang et al., 1995</td>
<td>CASE.B</td>
<td>N</td>
<td>-2</td>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>AD11</td>
<td>N</td>
<td>1</td>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>BABY A</td>
<td>N</td>
<td>1</td>
<td>25</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>HU</td>
<td>N</td>
<td>-1</td>
<td>32</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>X47</td>
<td>N</td>
<td>0</td>
<td>26</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>ACTG6</td>
<td>N</td>
<td>2</td>
<td>36</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>ACTG11</td>
<td>N</td>
<td>-1</td>
<td>31</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>X56</td>
<td>N</td>
<td>-1</td>
<td>32</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>MR</td>
<td>N</td>
<td>-2</td>
<td>29</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>X44</td>
<td>N</td>
<td>-5</td>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>AD6</td>
<td>N</td>
<td>-1</td>
<td>27</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>ACTG10</td>
<td>N</td>
<td>-1</td>
<td>36</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>ACTG8</td>
<td>N</td>
<td>0</td>
<td>29</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>EJ</td>
<td>N</td>
<td>-1</td>
<td>30</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>76B</td>
<td>N</td>
<td>0</td>
<td>26</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>CASE.C</td>
<td>N</td>
<td>-1</td>
<td>30</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>AD10</td>
<td>N</td>
<td>-2</td>
<td>29</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>AD13</td>
<td>N</td>
<td>-2</td>
<td>29</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>RT</td>
<td>N</td>
<td>1</td>
<td>32</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>JCS010</td>
<td>N</td>
<td>-3</td>
<td>32</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>CASE.D1</td>
<td>N</td>
<td>-1</td>
<td>35</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>CASE.D7</td>
<td>N</td>
<td>-3</td>
<td>27</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>AD8</td>
<td>N</td>
<td>0</td>
<td>39</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>N70</td>
<td>N</td>
<td>-1</td>
<td>42</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>277B</td>
<td>S</td>
<td>0</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>ACTG1</td>
<td>S</td>
<td>0</td>
<td>19</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>X57</td>
<td>S</td>
<td>-1</td>
<td>27</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>CASE.A2</td>
<td>S</td>
<td>-1</td>
<td>36</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>CASE.A1</td>
<td>S</td>
<td>-1</td>
<td>27</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>X42C</td>
<td>S</td>
<td>-3</td>
<td>28</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>ACTG3</td>
<td>S</td>
<td>-2</td>
<td>31</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>JSH</td>
<td>S</td>
<td>-2</td>
<td>29</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>X50C</td>
<td>S</td>
<td>0</td>
<td>37</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>X49</td>
<td>S</td>
<td>-2</td>
<td>48</td>
<td>8</td>
</tr>
<tr>
<td>Data Origin</td>
<td>Sequence</td>
<td>Phenotype</td>
<td>V2 Charge</td>
<td>V2 Length</td>
<td>V2 CHO Sites</td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>--------------</td>
</tr>
<tr>
<td>Wang et al., 1995</td>
<td>CASE.B</td>
<td>N</td>
<td>0</td>
<td>42</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>AD11</td>
<td>N</td>
<td>-1</td>
<td>44</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>BABYA</td>
<td>N</td>
<td>-1</td>
<td>42</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>HU</td>
<td>N</td>
<td>2</td>
<td>42</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>X47</td>
<td>N</td>
<td>1</td>
<td>44</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>ACTG6</td>
<td>N</td>
<td>1</td>
<td>41</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>ACTG11</td>
<td>N</td>
<td>1</td>
<td>41</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>X56</td>
<td>N</td>
<td>-1</td>
<td>41</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>MR</td>
<td>N</td>
<td>-1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>X44</td>
<td>N</td>
<td>-3</td>
<td>40</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>AD6</td>
<td>N</td>
<td>3</td>
<td>45</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ACTG10</td>
<td>N</td>
<td>0</td>
<td>43</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>ACTG8</td>
<td>N</td>
<td>2</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>EJ</td>
<td>N</td>
<td>-1</td>
<td>46</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>76B</td>
<td>N</td>
<td>3</td>
<td>46</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>CASE.C</td>
<td>N</td>
<td>0</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>AD10</td>
<td>N</td>
<td>0</td>
<td>43</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>AD13</td>
<td>N</td>
<td>2</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>RT</td>
<td>N</td>
<td>-2</td>
<td>42</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>JCS010</td>
<td>N</td>
<td>0</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>CASE.D1</td>
<td>N</td>
<td>1</td>
<td>47</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>CASE.D7</td>
<td>N</td>
<td>2</td>
<td>41</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>AD8</td>
<td>N</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>N70</td>
<td>N</td>
<td>2</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>277B</td>
<td>S</td>
<td>0</td>
<td>42</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>ACTG1</td>
<td>S</td>
<td>1</td>
<td>46</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>X57</td>
<td>S</td>
<td>0</td>
<td>41</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>CASE.A2</td>
<td>S</td>
<td>2</td>
<td>42</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>CASE.A1</td>
<td>S</td>
<td>4</td>
<td>43</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>X42C</td>
<td>S</td>
<td>0</td>
<td>41</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>ACTG3</td>
<td>S</td>
<td>2</td>
<td>41</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>JSH</td>
<td>S</td>
<td>1</td>
<td>46</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>X50C</td>
<td>S</td>
<td>-1</td>
<td>41</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>X49</td>
<td>S</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>Groenink et al., 1993</td>
<td>ACH-239.11</td>
<td>N</td>
<td>-1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Ama-44</td>
<td>N</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Ama-168.2</td>
<td>N</td>
<td>3</td>
<td>40</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Ama-96.1</td>
<td>N</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>ACH-424</td>
<td>N</td>
<td>0</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Ama-24</td>
<td>N</td>
<td>2</td>
<td>40</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ACH-15.9</td>
<td>N</td>
<td>-1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>ACH-172.1</td>
<td>N</td>
<td>3</td>
<td>42</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>ACH-525</td>
<td>N</td>
<td>-1</td>
<td>42</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>ACH-638</td>
<td>N</td>
<td>3</td>
<td>39</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Ama-179</td>
<td>N</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Ama-180</td>
<td>N</td>
<td>2</td>
<td>41</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Ama-181</td>
<td>N</td>
<td>3</td>
<td>39</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Ama-182</td>
<td>N</td>
<td>2</td>
<td>39</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Ama-161.15</td>
<td>S</td>
<td>1</td>
<td>49</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>ACH-320.2A.1.1</td>
<td>S</td>
<td>1</td>
<td>48</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Ama-16.2</td>
<td>S</td>
<td>3</td>
<td>43</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>ACH-39.14.2</td>
<td>S</td>
<td>1</td>
<td>47</td>
<td>3</td>
</tr>
<tr>
<td>Ama-72.3.3</td>
<td>S</td>
<td>0</td>
<td>42</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>ACH-224.25.5</td>
<td>S</td>
<td>2</td>
<td>41</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>ACH-571.16.2</td>
<td>S</td>
<td>1</td>
<td>42</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Ama-169.2</td>
<td>S</td>
<td>3</td>
<td>47</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>ACH-479.5</td>
<td>S</td>
<td>2</td>
<td>45</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>ACH-168.7</td>
<td>S</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Ama-55</td>
<td>S</td>
<td>2</td>
<td>43</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>ACH-704.2</td>
<td>S</td>
<td>-2</td>
<td>40</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ACH-373.38</td>
<td>S</td>
<td>0</td>
<td>40</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Ama-32</td>
<td>S</td>
<td>3</td>
<td>43</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Ama-127.4.2</td>
<td>S</td>
<td>3</td>
<td>42</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Cornelissen et al., 1995
<p>| BR03 | N | 1 | 38 | 2 |
| BR04 | N | 1 | 40 | 2 |
| BR17 | N | -1 | 43 | 3 |
| BR18c | N | 0 | 43 | 3 |
| BR19 | N | 0 | 39 | 2 |
| BR20c1 | N | -1 | 40 | 1 |
| BR20c2 | N | 0 | 38 | 1 |
| BR21 | N | 0 | 40 | 2 |
| BR23 | N | 3 | 40 | 2 |
| BR28c1 | N | 2 | 40 | 3 |
| BR28c2 | N | 0 | 40 | 2 |
| BR30 | N | 2 | 40 | 3 |
| TH14 | N | 0 | 39 | 2 |
| TH26 | N | 1 | 39 | 1 |
| AMC-01c | N | 0 | 40 | 1 |
| AMC-03 | N | 2 | 40 | 2 |
| AMC-04 | N | 1 | 50 | 4 |
| AMC-06 | N | 1 | 38 | 1 |
| AMC-12 | N | 1 | 42 | 2 |
| AMC-14 | N | 3 | 44 | 2 |
| AMC-15 | N | 0 | 46 | 3 |
| AMC-18 | N | 2 | 40 | 2 |
| BR14c | S | 2 | 38 | 1 |
| AMC-02 | S | 0 | 40 | 2 |
| AMC-05 | S | -1 | 61 | 5 |
| AMC-07c | S | 4 | 40 | 2 |
| AMC-8 | S | 2 | 39 | 1 |
| AMC-9 | S | 1 | 40 | 1 |
| AMC-10 | S | 0 | 45 | 1 |
| AMC-19 | S | 3 | 40 | 1 |
| AMC-20c1 | S | 0 | 41 | 2 |
| AMC-20c2 | S | -1 | 40 | 2 |
| AMC-21 | S | 4 | 40 | 2 |
| AMC-22 | S | 2 | 40 | 2 |
| AMC-24 | S | 0 | 38 | 1 |</p>
<table>
<thead>
<tr>
<th>Patient No.</th>
<th>Origin</th>
<th>V1 charge</th>
<th>V1 length</th>
<th>V1 CHO Sites</th>
<th>V2 charge</th>
<th>V2 length</th>
<th>V2 CHO Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 BR-1</td>
<td></td>
<td>-1</td>
<td>30</td>
<td>2</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 BR-2</td>
<td></td>
<td>-1</td>
<td>30</td>
<td>2</td>
<td>2</td>
<td>40</td>
<td>3</td>
</tr>
<tr>
<td>5 BR-3</td>
<td></td>
<td>-1</td>
<td>30</td>
<td>2</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 BR-4</td>
<td></td>
<td>0</td>
<td>30</td>
<td>3</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 BR-5</td>
<td></td>
<td>-2</td>
<td>30</td>
<td>3</td>
<td>-1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 BR-6</td>
<td></td>
<td>-1</td>
<td>30</td>
<td>2</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 BR-7</td>
<td></td>
<td>-1</td>
<td>30</td>
<td>2</td>
<td>2</td>
<td>40</td>
<td>3</td>
</tr>
<tr>
<td>5 BR-8</td>
<td></td>
<td>0</td>
<td>30</td>
<td>3</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 BR-9</td>
<td></td>
<td>-1</td>
<td>30</td>
<td>2</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 BR-10</td>
<td></td>
<td>-1</td>
<td>30</td>
<td>2</td>
<td>2</td>
<td>40</td>
<td>3</td>
</tr>
<tr>
<td>5 BR-11</td>
<td></td>
<td>-1</td>
<td>30</td>
<td>2</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 BR-12</td>
<td></td>
<td>0</td>
<td>30</td>
<td>3</td>
<td>2</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LN-1</td>
<td></td>
<td>2</td>
<td>30</td>
<td>3</td>
<td>2</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LN-2</td>
<td></td>
<td>-1</td>
<td>30</td>
<td>2</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LN-3</td>
<td></td>
<td>0</td>
<td>30</td>
<td>3</td>
<td>-1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LN-4</td>
<td></td>
<td>0</td>
<td>30</td>
<td>2</td>
<td>2</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LN-5</td>
<td></td>
<td>0</td>
<td>30</td>
<td>3</td>
<td>0</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LN-6</td>
<td></td>
<td>0</td>
<td>30</td>
<td>3</td>
<td>2</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LN-7</td>
<td></td>
<td>0</td>
<td>30</td>
<td>2</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LN-8</td>
<td></td>
<td>1</td>
<td>30</td>
<td>2</td>
<td>2</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LN-9</td>
<td></td>
<td>0</td>
<td>30</td>
<td>2</td>
<td>2</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LN-10</td>
<td></td>
<td>1</td>
<td>30</td>
<td>3</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LG-1</td>
<td></td>
<td>-1</td>
<td>30</td>
<td>3</td>
<td>0</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LG-2</td>
<td></td>
<td>0</td>
<td>30</td>
<td>3</td>
<td>0</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LG-3</td>
<td></td>
<td>0</td>
<td>30</td>
<td>3</td>
<td>0</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LG-4</td>
<td></td>
<td>0</td>
<td>30</td>
<td>3</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LG-5</td>
<td></td>
<td>-1</td>
<td>30</td>
<td>3</td>
<td>0</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LG-6</td>
<td></td>
<td>-1</td>
<td>30</td>
<td>3</td>
<td>0</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LG-7</td>
<td></td>
<td>0</td>
<td>30</td>
<td>3</td>
<td>0</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LG-8</td>
<td></td>
<td>1</td>
<td>30</td>
<td>3</td>
<td>0</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LG-9</td>
<td></td>
<td>0</td>
<td>30</td>
<td>3</td>
<td>-1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LG-10</td>
<td></td>
<td>1</td>
<td>30</td>
<td>3</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LG-11</td>
<td></td>
<td>0</td>
<td>30</td>
<td>3</td>
<td>0</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LG-12</td>
<td></td>
<td>1</td>
<td>30</td>
<td>3</td>
<td>3</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LG-13</td>
<td></td>
<td>0</td>
<td>30</td>
<td>3</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LG-14</td>
<td></td>
<td>0</td>
<td>30</td>
<td>3</td>
<td>2</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>5 LG-15</td>
<td></td>
<td>0</td>
<td>30</td>
<td>3</td>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>9 BR-1</td>
<td></td>
<td>2</td>
<td>31</td>
<td>2</td>
<td>0</td>
<td>43</td>
<td>2</td>
</tr>
<tr>
<td>9 BR-2</td>
<td></td>
<td>1</td>
<td>30</td>
<td>4</td>
<td>2</td>
<td>43</td>
<td>2</td>
</tr>
<tr>
<td>9 BR-3</td>
<td></td>
<td>2</td>
<td>32</td>
<td>4</td>
<td>1</td>
<td>43</td>
<td>2</td>
</tr>
<tr>
<td>9 BR-4</td>
<td></td>
<td>-1</td>
<td>31</td>
<td>4</td>
<td>2</td>
<td>43</td>
<td>2</td>
</tr>
<tr>
<td>9 BR-5</td>
<td></td>
<td>0</td>
<td>35</td>
<td>4</td>
<td>2</td>
<td>43</td>
<td>2</td>
</tr>
<tr>
<td>9 BR-6</td>
<td></td>
<td>0</td>
<td>32</td>
<td>4</td>
<td>1</td>
<td>43</td>
<td>2</td>
</tr>
<tr>
<td>9 BR-7</td>
<td></td>
<td>0</td>
<td>32</td>
<td>4</td>
<td>3</td>
<td>43</td>
<td>2</td>
</tr>
<tr>
<td>9 BR-8</td>
<td></td>
<td>1</td>
<td>31</td>
<td>4</td>
<td>3</td>
<td>44</td>
<td>3</td>
</tr>
<tr>
<td>9 BR-9</td>
<td></td>
<td>2</td>
<td>27</td>
<td>4</td>
<td>1</td>
<td>48</td>
<td>4</td>
</tr>
<tr>
<td>9 BR-10</td>
<td></td>
<td>3</td>
<td>38</td>
<td>5</td>
<td>3</td>
<td>43</td>
<td>2</td>
</tr>
<tr>
<td>9 BR-11</td>
<td></td>
<td>0</td>
<td>38</td>
<td>5</td>
<td>2</td>
<td>43</td>
<td>2</td>
</tr>
<tr>
<td>9 LN-1</td>
<td></td>
<td>0</td>
<td>38</td>
<td>5</td>
<td>0</td>
<td>47</td>
<td>5</td>
</tr>
<tr>
<td>9 LN-2</td>
<td></td>
<td>1</td>
<td>31</td>
<td>5</td>
<td>0</td>
<td>47</td>
<td>5</td>
</tr>
<tr>
<td>9 LN-3</td>
<td></td>
<td>0</td>
<td>38</td>
<td>5</td>
<td>-1</td>
<td>47</td>
<td>5</td>
</tr>
<tr>
<td>9 LN-4</td>
<td></td>
<td>1</td>
<td>31</td>
<td>4</td>
<td>0</td>
<td>47</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient</td>
<td>Disease Status</td>
<td>CD4 count</td>
<td>V1 No. LN</td>
<td>V1 No. BR</td>
<td>V2 No. LN</td>
<td>V2 No. BR</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>4 S</td>
<td></td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5 S</td>
<td></td>
<td>95</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>6 S</td>
<td></td>
<td>10</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>15 S</td>
<td></td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>21 S</td>
<td></td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>26 S</td>
<td></td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>30 S</td>
<td></td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>32 S</td>
<td></td>
<td>11</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>42 S</td>
<td></td>
<td>28</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>44 S</td>
<td></td>
<td>19</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>51 S</td>
<td></td>
<td>10</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>63 S</td>
<td></td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>69 S</td>
<td></td>
<td>90</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>10 S</td>
<td></td>
<td>6</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 S</td>
<td></td>
<td>4</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 S</td>
<td></td>
<td>12</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 S</td>
<td></td>
<td>10</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 S</td>
<td></td>
<td>19</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 S</td>
<td></td>
<td>93</td>
<td>6</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 S</td>
<td></td>
<td>5</td>
<td>7</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28 S</td>
<td></td>
<td>105</td>
<td>10</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36 S</td>
<td></td>
<td>15</td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37 S</td>
<td></td>
<td>2</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38 S</td>
<td></td>
<td>40</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39 S</td>
<td></td>
<td>170</td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45 S</td>
<td></td>
<td>5</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48 S</td>
<td></td>
<td>40</td>
<td>5</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54 S</td>
<td></td>
<td>13</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62 S</td>
<td></td>
<td>2</td>
<td>2</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64 S</td>
<td></td>
<td>2</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68 S</td>
<td></td>
<td>155</td>
<td>7</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>72 S</td>
<td></td>
<td>140</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>77 S</td>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>78 S</td>
<td></td>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 P</td>
<td></td>
<td>115</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 P</td>
<td></td>
<td>140</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 P</td>
<td></td>
<td>370</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 P</td>
<td></td>
<td>33</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34 P</td>
<td></td>
<td>240</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35 P</td>
<td></td>
<td>180</td>
<td>6</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 P</td>
<td></td>
<td>50</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76 P</td>
<td></td>
<td>802</td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient</td>
<td>Disease status</td>
<td>Tissue origin</td>
<td>Actual length</td>
<td>Genomic region</td>
<td>CD4 count</td>
<td>GCE</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
<td>---------------</td>
<td>---------------</td>
<td>----------------</td>
<td>-----------</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>4 S</td>
<td>LN</td>
<td>30 V1</td>
<td>8 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 S</td>
<td>BR</td>
<td>30 V1</td>
<td>8 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 S</td>
<td>LN</td>
<td>31 V1</td>
<td>95 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 S</td>
<td>LN</td>
<td>32 V1</td>
<td>95 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 S</td>
<td>LN</td>
<td>36 V1</td>
<td>95 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 S</td>
<td>LN</td>
<td>38 V1</td>
<td>95 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 S</td>
<td>LN</td>
<td>40 V1</td>
<td>95 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 S</td>
<td>BR</td>
<td>31 V1</td>
<td>95 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 S</td>
<td>BR</td>
<td>38 V1</td>
<td>95 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 S</td>
<td>BR</td>
<td>40 V1</td>
<td>95 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 S</td>
<td>LN</td>
<td>30 V1</td>
<td>10 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 S</td>
<td>LN</td>
<td>31 V1</td>
<td>10 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 S</td>
<td>LN</td>
<td>32 V1</td>
<td>10 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 S</td>
<td>LN</td>
<td>40 V1</td>
<td>10 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 S</td>
<td>LN</td>
<td>41 V1</td>
<td>10 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 S</td>
<td>LN</td>
<td>42 V1</td>
<td>10 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 S</td>
<td>BR</td>
<td>30 V1</td>
<td>10 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 S</td>
<td>BR</td>
<td>32 V1</td>
<td>10 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 S</td>
<td>BR</td>
<td>42 V1</td>
<td>10 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 S</td>
<td>LN</td>
<td>30 V1</td>
<td>5 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 S</td>
<td>BR</td>
<td>30 V1</td>
<td>5 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 S</td>
<td>BR</td>
<td>36 V1</td>
<td>5 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 S</td>
<td>LN</td>
<td>35 V1</td>
<td>6 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 S</td>
<td>LN</td>
<td>39 V1</td>
<td>6 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 S</td>
<td>BR</td>
<td>31 V1</td>
<td>6 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 S</td>
<td>BR</td>
<td>35 V1</td>
<td>6 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 S</td>
<td>BR</td>
<td>42 V1</td>
<td>6 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 S</td>
<td>LN</td>
<td>31 V1</td>
<td>8 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 S</td>
<td>LN</td>
<td>32 V1</td>
<td>8 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 S</td>
<td>LN</td>
<td>39 V1</td>
<td>8 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 S</td>
<td>LN</td>
<td>49 V1</td>
<td>8 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 S</td>
<td>LN</td>
<td>50 V1</td>
<td>8 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 S</td>
<td>BR</td>
<td>31 V1</td>
<td>8 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 S</td>
<td>BR</td>
<td>32 V1</td>
<td>8 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 S</td>
<td>BR</td>
<td>39 V1</td>
<td>8 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 S</td>
<td>LN</td>
<td>31 V1</td>
<td>0 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 S</td>
<td>LN</td>
<td>39 V1</td>
<td>0 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 S</td>
<td>BR</td>
<td>31 V1</td>
<td>0 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 S</td>
<td>BR</td>
<td>39 V1</td>
<td>0 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 S</td>
<td>LN</td>
<td>28 V1</td>
<td>11 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 S</td>
<td>LN</td>
<td>32 V1</td>
<td>11 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 S</td>
<td>LN</td>
<td>36 V1</td>
<td>11 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 S</td>
<td>LN</td>
<td>37 V1</td>
<td>11 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 S</td>
<td>LN</td>
<td>39 V1</td>
<td>11 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 S</td>
<td>BR</td>
<td>28 V1</td>
<td>11 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 S</td>
<td>BR</td>
<td>37 V1</td>
<td>11 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34 P</td>
<td>LN</td>
<td>30 V1</td>
<td>240 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 S</td>
<td>LN</td>
<td>31 V1</td>
<td>50 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 S</td>
<td>LN</td>
<td>33 V1</td>
<td>50 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 S</td>
<td>LN</td>
<td>39 V1</td>
<td>50 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 S</td>
<td>BR</td>
<td>31 V1</td>
<td>50 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 S</td>
<td>BR</td>
<td>39 V1</td>
<td>50 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 S</td>
<td>BR</td>
<td>40 V1</td>
<td>50 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>----</td>
<td>-------</td>
<td>------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42 S</td>
<td>LN</td>
<td>27 V1</td>
<td>28 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42 S</td>
<td>LN</td>
<td>30 V1</td>
<td>28 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42 S</td>
<td>LN</td>
<td>34 V1</td>
<td>28 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42 S</td>
<td>BR</td>
<td>27 V1</td>
<td>28 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42 S</td>
<td>BR</td>
<td>34 V1</td>
<td>28 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42 S</td>
<td>BR</td>
<td>37 V1</td>
<td>28 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44 S</td>
<td>LN</td>
<td>27 V1</td>
<td>19 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44 S</td>
<td>LN</td>
<td>34 V1</td>
<td>19 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>63 S</td>
<td>LN</td>
<td>31 V1</td>
<td>2 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>63 S</td>
<td>LN</td>
<td>31 V1</td>
<td>2 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>63 S</td>
<td>BR</td>
<td>31 V1</td>
<td>2 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>63 S</td>
<td>BR</td>
<td>39 V1</td>
<td>2 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>69 S</td>
<td>LN</td>
<td>22 V1</td>
<td>90 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>69 S</td>
<td>LN</td>
<td>29 V1</td>
<td>90 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>69 S</td>
<td>BR</td>
<td>27 V1</td>
<td>90 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>69 S</td>
<td>BR</td>
<td>30 V1</td>
<td>90 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>69 S</td>
<td>BR</td>
<td>38 V1</td>
<td>90 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>77 S</td>
<td>LN</td>
<td>26 V1</td>
<td>3 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>77 S</td>
<td>LN</td>
<td>27 V1</td>
<td>3 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>77 S</td>
<td>LN</td>
<td>33 V1</td>
<td>3 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>78 S</td>
<td>LN</td>
<td>36 V1</td>
<td>0 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>78 S</td>
<td>LN</td>
<td>46 V1</td>
<td>0 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 S</td>
<td>LN</td>
<td>37 V1</td>
<td>6 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 S</td>
<td>LN</td>
<td>47 V1</td>
<td>6 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 P</td>
<td>LN</td>
<td>30 V1</td>
<td>115 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 P</td>
<td>LN</td>
<td>31 V1</td>
<td>115 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 P</td>
<td>LN</td>
<td>38 V1</td>
<td>115 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 P</td>
<td>LN</td>
<td>39 V1</td>
<td>115 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 P</td>
<td>LN</td>
<td>30 V1</td>
<td>370 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 P</td>
<td>LN</td>
<td>31 V1</td>
<td>370 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 P</td>
<td>LN</td>
<td>38 V1</td>
<td>370 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 S</td>
<td>LN</td>
<td>34 V1</td>
<td>4 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 S</td>
<td>LN</td>
<td>38 V1</td>
<td>4 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 S</td>
<td>LN</td>
<td>41 V1</td>
<td>4 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 S</td>
<td>LN</td>
<td>36 V1</td>
<td>12 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 S</td>
<td>LN</td>
<td>42 V1</td>
<td>12 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 S</td>
<td>LN</td>
<td>24 V1</td>
<td>10 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 S</td>
<td>LN</td>
<td>25 V1</td>
<td>10 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 S</td>
<td>LN</td>
<td>30 V1</td>
<td>10 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 S</td>
<td>LN</td>
<td>32 V1</td>
<td>10 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 P</td>
<td>LN</td>
<td>27 V1</td>
<td>33 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 P</td>
<td>LN</td>
<td>36 V1</td>
<td>33 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 P</td>
<td>LN</td>
<td>29 V1</td>
<td>140 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 P</td>
<td>LN</td>
<td>34 V1</td>
<td>140 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 P</td>
<td>LN</td>
<td>39 V1</td>
<td>140 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 P</td>
<td>LN</td>
<td>45 V1</td>
<td>140 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 S</td>
<td>LN</td>
<td>26 V1</td>
<td>19 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 S</td>
<td>LN</td>
<td>27 V1</td>
<td>19 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 S</td>
<td>LN</td>
<td>29 V1</td>
<td>19 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 S</td>
<td>LN</td>
<td>36 V1</td>
<td>19 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 P</td>
<td>LN</td>
<td>37 V1</td>
<td>93 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 P</td>
<td>LN</td>
<td>38 V1</td>
<td>93 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 P</td>
<td>LN</td>
<td>40 V1</td>
<td>93 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 P</td>
<td>LN</td>
<td>43 V1</td>
<td>93 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 P</td>
<td>LN</td>
<td>44 V1</td>
<td>93 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 S</td>
<td>LN</td>
<td>25 V1</td>
<td>5 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 S</td>
<td>LN</td>
<td>27 V1</td>
<td>5 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 S</td>
<td>LN</td>
<td>28 V1</td>
<td>5 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 S</td>
<td>LN</td>
<td>32 V1</td>
<td>5 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 S</td>
<td>LN</td>
<td>38 V1</td>
<td>5 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 S</td>
<td>LN</td>
<td>40 V1</td>
<td>5 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35 P</td>
<td>LN</td>
<td>29 V1</td>
<td>180 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35 P</td>
<td>LN</td>
<td>31 V1</td>
<td>180 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35 P</td>
<td>LN</td>
<td>33 V1</td>
<td>180 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35 P</td>
<td>LN</td>
<td>36 V1</td>
<td>180 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35 P</td>
<td>LN</td>
<td>37 V1</td>
<td>180 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35 P</td>
<td>LN</td>
<td>38 V1</td>
<td>180 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36 S</td>
<td>LN</td>
<td>30 V1</td>
<td>15 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36 S</td>
<td>LN</td>
<td>36 V1</td>
<td>15 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37 S</td>
<td>LN</td>
<td>35 V1</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37 S</td>
<td>LN</td>
<td>36 V1</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37 S</td>
<td>LN</td>
<td>43 V1</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37 S</td>
<td>LN</td>
<td>44 V1</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38 S</td>
<td>LN</td>
<td>24 V1</td>
<td>40 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38 S</td>
<td>LN</td>
<td>30 V1</td>
<td>40 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38 S</td>
<td>LN</td>
<td>37 V1</td>
<td>40 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39 S</td>
<td>LN</td>
<td>36 V1</td>
<td>170 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39 S</td>
<td>LN</td>
<td>38 V1</td>
<td>170 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45 S</td>
<td>LN</td>
<td>30 V1</td>
<td>5 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45 S</td>
<td>LN</td>
<td>37 V1</td>
<td>5 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48 S</td>
<td>LN</td>
<td>28 V1</td>
<td>40 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48 S</td>
<td>LN</td>
<td>31 V1</td>
<td>40 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48 S</td>
<td>LN</td>
<td>35 V1</td>
<td>40 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48 S</td>
<td>LN</td>
<td>38 V1</td>
<td>40 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48 S</td>
<td>LN</td>
<td>39 V1</td>
<td>40 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54 S</td>
<td>LN</td>
<td>29 V1</td>
<td>3 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54 S</td>
<td>LN</td>
<td>36 V1</td>
<td>3 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54 S</td>
<td>LN</td>
<td>39 V1</td>
<td>3 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54 S</td>
<td>LN</td>
<td>43 V1</td>
<td>3 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62 S</td>
<td>LN</td>
<td>30 V1</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62 S</td>
<td>LN</td>
<td>37 V1</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64 S</td>
<td>LN</td>
<td>36 V1</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64 S</td>
<td>LN</td>
<td>37 V1</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64 S</td>
<td>LN</td>
<td>44 V1</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68 S</td>
<td>LN</td>
<td>35 V1</td>
<td>155 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68 S</td>
<td>LN</td>
<td>36 V1</td>
<td>155 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68 S</td>
<td>LN</td>
<td>39 V1</td>
<td>155 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68 S</td>
<td>LN</td>
<td>41 V1</td>
<td>155 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68 S</td>
<td>LN</td>
<td>42 V1</td>
<td>155 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68 S</td>
<td>LN</td>
<td>43 V1</td>
<td>155 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68 S</td>
<td>LN</td>
<td>44 V1</td>
<td>155 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>72 S</td>
<td>LN</td>
<td>27 V1</td>
<td>140 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>72 S</td>
<td>LN</td>
<td>35 V1</td>
<td>140 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76 P</td>
<td>LN</td>
<td>43 V1</td>
<td>802 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 S</td>
<td>BR</td>
<td>40 V2</td>
<td>8 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 S</td>
<td>BR</td>
<td>40 V2</td>
<td>8 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>LN</td>
<td>V2</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>----</td>
<td>----</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>LN</td>
<td>41</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>LN</td>
<td>43</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>LN</td>
<td>44</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>LN</td>
<td>47</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>BR</td>
<td>43</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>BR</td>
<td>44</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>BR</td>
<td>48</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>LN</td>
<td>38</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>LN</td>
<td>40</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>LN</td>
<td>41</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>BR</td>
<td>38</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>BR</td>
<td>41</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>LN</td>
<td>41</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>LN</td>
<td>42</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>LN</td>
<td>48</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>BR</td>
<td>46</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>BR</td>
<td>48</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>LN</td>
<td>41</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>BR</td>
<td>41</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>LN</td>
<td>43</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>BR</td>
<td>43</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>LN</td>
<td>37</td>
<td>240</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>LN</td>
<td>40</td>
<td>240</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>LN</td>
<td>40</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>LN</td>
<td>38</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>LN</td>
<td>38</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>LN</td>
<td>40</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>BR</td>
<td>40</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>LN</td>
<td>41</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>LN</td>
<td>43</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>LN</td>
<td>44</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>LN</td>
<td>39</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>LN</td>
<td>43</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>BR</td>
<td>39</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>BR</td>
<td>43</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>LN</td>
<td>37</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>LN</td>
<td>40</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>BR</td>
<td>40</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>LN</td>
<td>40</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>LN</td>
<td>42</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>BR</td>
<td>40</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>LN</td>
<td>43</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>LN</td>
<td>44</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>LN</td>
<td>45</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>LN</td>
<td>42</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>LN</td>
<td>45</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>LN</td>
<td>41</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>LN</td>
<td>44</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>LN</td>
<td>41</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>LN</td>
<td>39</td>
<td>115</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>LN</td>
<td>39</td>
<td>115</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>LN</td>
<td>41</td>
<td>115</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 P</td>
<td>LN</td>
<td>43 V2</td>
<td>115 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 P</td>
<td>LN</td>
<td>41 V2</td>
<td>370 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 P</td>
<td>LN</td>
<td>49 V2</td>
<td>370 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 S</td>
<td>LN</td>
<td>36 V2</td>
<td>4 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 S</td>
<td>LN</td>
<td>37 V2</td>
<td>4 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 S</td>
<td>LN</td>
<td>40 V2</td>
<td>4 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 S</td>
<td>LN</td>
<td>37 V2</td>
<td>12 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 S</td>
<td>LN</td>
<td>38 V2</td>
<td>12 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 S</td>
<td>LN</td>
<td>41 V2</td>
<td>12 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 S</td>
<td>LN</td>
<td>35 V2</td>
<td>10 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 S</td>
<td>LN</td>
<td>36 V2</td>
<td>10 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 S</td>
<td>LN</td>
<td>38 V2</td>
<td>10 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 P</td>
<td>LN</td>
<td>37 V2</td>
<td>33 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 P</td>
<td>LN</td>
<td>40 V2</td>
<td>33 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 P</td>
<td>LN</td>
<td>37 V2</td>
<td>140 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 P</td>
<td>LN</td>
<td>40 V2</td>
<td>140 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 S</td>
<td>LN</td>
<td>37 V2</td>
<td>19 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 S</td>
<td>LN</td>
<td>39 V2</td>
<td>19 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 S</td>
<td>LN</td>
<td>40 V2</td>
<td>19 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 S</td>
<td>LN</td>
<td>44 V2</td>
<td>19 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 P</td>
<td>LN</td>
<td>38 V2</td>
<td>93 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 P</td>
<td>LN</td>
<td>40 V2</td>
<td>93 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 S</td>
<td>LN</td>
<td>38 V2</td>
<td>5 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 S</td>
<td>LN</td>
<td>40 V2</td>
<td>5 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35 P</td>
<td>LN</td>
<td>46 V2</td>
<td>180 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35 P</td>
<td>LN</td>
<td>49 V2</td>
<td>180 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36 S</td>
<td>LN</td>
<td>38 V2</td>
<td>15 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36 S</td>
<td>LN</td>
<td>39 V2</td>
<td>15 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36 S</td>
<td>LN</td>
<td>40 V2</td>
<td>15 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36 S</td>
<td>LN</td>
<td>41 V2</td>
<td>15 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36 S</td>
<td>LN</td>
<td>43 V2</td>
<td>15 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37 S</td>
<td>LN</td>
<td>38 V2</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37 S</td>
<td>LN</td>
<td>40 V2</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38 S</td>
<td>LN</td>
<td>40 V2</td>
<td>40 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38 S</td>
<td>LN</td>
<td>42 V2</td>
<td>40 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39 S</td>
<td>LN</td>
<td>38 V2</td>
<td>170 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39 S</td>
<td>LN</td>
<td>40 V2</td>
<td>170 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39 S</td>
<td>LN</td>
<td>41 V2</td>
<td>170 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39 S</td>
<td>LN</td>
<td>43 V2</td>
<td>170 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45 S</td>
<td>LN</td>
<td>34 V2</td>
<td>5 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45 S</td>
<td>LN</td>
<td>37 V2</td>
<td>5 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45 S</td>
<td>LN</td>
<td>39 V2</td>
<td>5 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48 S</td>
<td>LN</td>
<td>36 V2</td>
<td>40 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48 S</td>
<td>LN</td>
<td>39 V2</td>
<td>40 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54 S</td>
<td>LN</td>
<td>34 V2</td>
<td>3 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54 S</td>
<td>LN</td>
<td>36 V2</td>
<td>3 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54 S</td>
<td>LN</td>
<td>38 V2</td>
<td>3 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54 S</td>
<td>LN</td>
<td>41 V2</td>
<td>3 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62 S</td>
<td>LN</td>
<td>34 V2</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62 S</td>
<td>LN</td>
<td>36 V2</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62 S</td>
<td>LN</td>
<td>38 V2</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62 S</td>
<td>LN</td>
<td>41 V2</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62 S</td>
<td>LN</td>
<td>43 V2</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64 S</td>
<td>LN</td>
<td>35 V2</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>----</td>
<td>-------</td>
<td>-----</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64 S</td>
<td>LN</td>
<td>37 V2</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64 S</td>
<td>LN</td>
<td>42 V2</td>
<td>2 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68 S</td>
<td>LN</td>
<td>34 V2</td>
<td>155 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68 S</td>
<td>LN</td>
<td>35 V2</td>
<td>155 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68 S</td>
<td>LN</td>
<td>37 V2</td>
<td>155 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68 S</td>
<td>LN</td>
<td>40 V2</td>
<td>155 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68 S</td>
<td>LN</td>
<td>42 V2</td>
<td>155 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>72 S</td>
<td>LN</td>
<td>34 V2</td>
<td>140 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>72 S</td>
<td>LN</td>
<td>37 V2</td>
<td>140 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>72 S</td>
<td>LN</td>
<td>40 V2</td>
<td>140 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76 P</td>
<td>LN</td>
<td>36 V2</td>
<td>802 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76 P</td>
<td>LN</td>
<td>37 V2</td>
<td>802 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76 P</td>
<td>LN</td>
<td>39 V2</td>
<td>802 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76 P</td>
<td>LN</td>
<td>42 V2</td>
<td>802 O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Investigation of the Dynamics of the Spread of Human Immuno deficiency Virus to Brain and Other Tissues by Evolutionary Analysis of Sequences from the p17gag and env Genes

E. S. HUGHES, J. E. BELL, AND P. SIMMONDS

Department of Medical Microbiology, University of Edinburgh, Edinburgh EH8 9AG, and Department of Neuropathology, University of Edinburgh, Western General Hospital, Edinburgh EH4 2UX, United Kingdom

Received 22 May 1996/Accepted 15 October 1996

The time of spread of human immunodeficiency virus type 1 (HIV-1) from lymphoid to nonlymphoid tissues in the course of infection was investigated by sequence comparisons of variants infecting a range of lymphoid and nonlymphoid tissues from three individuals with AIDS in the p17gag gene and regions flanking the V1/V2 hypervariable regions. Phylogenetic analysis in both regions revealed several lineages in each individual that contained sequences from both lymphoid and nonlymphoid tissues such as the brain. This observation contrasts strongly with the previously described organ-specific sequences in the V3 region in this study population and other investigations. Although individual pairwise comparisons of relatively short sequences such as p17gag are subject to considerable stochastic error, we found that the diversity of gag sequences in variants from lymphoid tissue was consistently lower than that found among variants amplified from the brain. By estimating mean synonymous pairwise distances in the p17gag region, we were able to make an approximate calculation of the ages of populations in different tissues. Those from lymphoid tissue ranged from 2.65 to 5.6 years in the three study subjects, compared with 4.1 to 6.2 years for variants in the brain. Indeed, variants infecting the brain were no more closely related to each other than they were to variants infecting other tissues in the body. In two of the three individuals, these times of divergence indicate that infection of the brain may have occurred as an early event in the progression to disease, preceding the onset of AIDS by several years. This study is the first in which it was possible to estimate times of diversification in different tissues in vivo and is of importance in understanding the dynamics of the spread of HIV-1 into nonlymphoid tissues and its possible adaptation for replication in different cell types.

Infection with human immunodeficiency virus (HIV) is associated with a slow, progressive, and irreversible impairment of the immune system, eventually leading to AIDS. Inherent in the nature of infection with HIV type 1 (HIV-1) is the prolonged asymptomatic period that precedes the development of disease (2, 11, 28, 35), where infection may be subclinical for as long as 10 to 15 years. This phenomenon was originally hypothesized to result from viral latency, whereby viral or proviral DNA became integrated into the host genome with the simultaneous cessation of viral expression and independent replication (2). The ensuing progression to AIDS would then result from subsequent reactivation of virus replication by various factors acting upon infected cells, such as antigens, mitogens, and transcriptional factors produced by other viruses. However, it has been recently shown that from the time of seroconversion, there is active replication of the virus in lymphoid tissues (11, 28, 35). There are few convincing demonstrations of active infection of nonlymphoid tissues until later in infection, and this change in distribution may be associated with increased immunosuppression in AIDS (9). Alternatively, it is possible that variants detected in nonlymphoid tissues such as the brain in patients with AIDS have been continuously present from initial infection but that infection becomes clinically significant only during severe immunosuppression. In this model, HIV encephalitis could be regarded as reactivation rather than de novo infection.

This study was undertaken to estimate the time of spread of HIV-1 to nonlymphoid tissues to determine whether reactivation or actual virus spread was responsible for the pathology observed in nonlymphoid tissues in AIDS. In order to do this, we obtained sequences from the p17gag region and V1/V2 flanking regions of HIV-1. The p17gag region was chosen because most nucleotide differences in this region are synonymous and therefore are not subject to positive selection pressures for sequence change, such as those that may be encountered by immunological recognition by antibody or cytotoxic T cells (36). Variations at silent sites occur at frequencies similar to those in the rest of the genome, and it has already been demonstrated in previous epidemiological studies that sequence relationships in this region reflect the evolutionary history of the virus (15, 17, 20). The rate of sequence change of the p17gag region has previously been determined from hemophilias infected from a common source (20), allowing the time of divergence between any pairs of sequences to be determined. This region is therefore of use in reconstructing epidemiological relationships between HIV-1-infected individuals (17) and can be extended to the comparison of variants within different cell types within a single infected individual.

Tissues from various lymphoid and nonlymphoid organs were obtained at autopsy from a number of HIV-1-positive patients known to have a high viral load in the brain and evidence of giant cell encephalitis by pathology. Phylogenetic analyses of both p17gag and V1/V2 flanking regions were carried out in order to explore the relationships among the various lineages present and the spread of infection to nonlymphoid tissues. It was possible to estimate the time of divergence...
between lymphoid and nonlymphoid tissues, allowing us to determine the length of time prior to death at which nonlymphoid tissues become infected.

MATERIALS AND METHODS

Patient samples. Tissues from various organs were obtained at autopsy, carried out within 3 days of death, from three individuals who died with AIDS-defining illnesses. All individuals showed evidence of HIV infection of the brain upon postmortem examination, as determined by the histological appearance of giant cells, the detection of p24 by immunochemistry, and the finding of high levels of HIV DNA in the brain by quantitative PCR (9). Pathological examination of the fixed brains revealed evidence of atrophy on external examination, and this was confirmed on section by the presence of ventricular dilatation and opening up of the sulci in all three patients. In one case (P6), a focal 1-cm-diameter lesion was identified on macroscopic inspection in the right basal ganglia. Histological examination of this lesion showed that it was a primary central nervous system lymphoma. Neither of the other two patients showed macroscopic focal lesions of the brain. Histological examination in all three patients displayed evidence of quite florid HIV encephalitis and leukoencephalopathy, characterized by giant cells and focal collections of macrophages and microglial cells, associated with myelin damage. There was no evidence of perivascular or leptomeningeal inflammatory infiltrates, and, in particular, lymphocytes were not identified within the central nervous system parenchyma. Further results of the pathological examinations of these three individuals and quantitation of HIV DNA sequences in brain and other tissues have been reported previously (8, 9). Clinical information in addition to the previous description (9) for the 4 or 5 years prior to death is summarized in Fig. 1. Samples of brain (left frontal lobe), spinal cord, lymph node (mesenteric), lung, and colon tissues from these patients were dissected into 1- to 2-cm pieces and stored at -70°C.

Preparation of DNA. Extraction of DNA from these tissues was carried out by resuspending small pieces of tissue in 500 μl of lysis buffer (50 mM Tris hydrochloride [pH 8.0], 100 mM NaCl, 50 mM EDTA, 1% sodium-dodecyl sulfate, 100 μg of proteinase K per ml). The digestion process was allowed to continue for 2 h at 65°C. This was followed by phenol-chloroform extraction and ethanol precipitation. DNA pellets were dried and resuspended in 100 to 200 μl of distilled water. The concentration of DNA in each sample was determined by UV absorbance at wavelengths of 260 and 280 nm.

Detection of provirus. Proviral DNA was amplified and quantified by a previously described limiting dilution and nested PCR method (45). Amplification of DNA was carried out with primers flanking hypervariable regions 1 and 2 from env and p17 from gag. The nucleotide sequences of the primers were as follows: for V1/V2, a, GAG GAT ATA ATG TTA TGG, + (sense), 6539; b, GA TCA AAG CCT AAA GCC ATG, +, 6560; c, TTG AAA GAG CAG TTT, − (antisense), 6677; d, TGT AA CAA AAA ACT CCT TTA CAA, +, 6684; e, CAA TAA TGG ATG GAA ATT GG, −, 6844; and f, AAT GAT CCA CCT TCA GG, +, 705. Sufficient DNA was amplified by using a thermal cycle of 36 s at 94°C, 42 s at 50°C for gag or 46°C for V1/V2, and 40 s at 72°C for strand extension. Each template strand was subjected to 25 cycles of amplification.

Sequence analysis. Single molecules of HIV provirus were isolated by limiting dilution and amplified in a nested PCR to produce sufficient DNA to allow direct sequencing of the PCR products. Direct sequencing of amplified DNA was achieved by using a solid-phase sequencing method. The second PCR was performed in a 100-μl volume with one biotin-labeled primer and one unlabeled primer (5 to 10 pmol per reaction mixture), generating a PCR product with one strand having a biotin moiety at either the 5′ or 3′ end. PCR products were immobilized on streptavidin-coated magnetic beads (Dynal), and single strands of DNA were purified by magnetic separation and sequenced according to the manufacturer's (Sequenase version 2.0) protocol. After this sequencing reaction, 5 to 6 μl of the reaction product was then electrophoresed on a denaturing polyacrylamide gel (6% acrylamide, 0.3% N,N-bisacrylamide). 8 M urea, 0.089 M Tris, 0.089 M boric acid, 0.002 M EDTA, pH 8.3). Gels were fixed, dried, and exposed overnight on X-Omat film.

Phylogenetic analysis. Sequence comparisons between viruses from the three study patients were made for the pl17<sup>env</sup> gene and hypervariable flanking regions of V1 and V2 of the env gene of HIV-1. The amplified pl17<sup>env</sup> region began at nucleotide 795 of HXB2 and extended to position 1319. The amplified V1/V2 region began at position 6539 of HXB2 and extended to position 6976. The length of the gag region used for sequence comparisons was 415 nucleotides, and that of the V1/V2 region was 297 nucleotides. An unrooted phylogenetic tree for all 85 pl17<sup>env</sup> nucleotide sequences and 99 V1/V2 nucleotide sequences obtained from lymph node, brain, and lung samples was constructed by the neighbor-joining method using the NEIGHBOR program in the PHYLIP package (version 3.5) (12). Distances between each pair of sequences were estimated by using the DNADIST program in the PHYLIP package (version 3.5) (15). Rooted trees were constructed for each patient by bootstrap resampling (500 replications) using the MEGA package with the sequence of HIVMN as an outgroup (25). Phylogenetic analysis of the env region was confined to regions flanking the V1 and V2 hypervariable regions because of the indeterminance and often arbitrary alignment of the hypervariable sequences.

Nucleotide sequence accession numbers. The sequences obtained in this study have been submitted to GenBank and assigned accession numbers U79869 (gag) and U79870 to U79871 and U79872 (U1/U2).

RESULTS

Rate of sequence change in pl17<sup>env</sup> regions. An unrooted neighbor-joining tree was constructed by using 85 sequences from the pl17<sup>env</sup> region (positions 835 to 1270 in the HXB2 clone [33]) from a range of lymphoid and nonlymphoid tissues of three HIV-infected individuals dying from AIDS. The sequences from each of the three study patients were distinct, grouping separately into three clusters. Bootstrap resampling supported the distinction of three separate groups (Fig. 2).

All three study subjects were infected with HIV through drug abuse in 1982 or 1983. Previous phylogenetic studies have implicated a common source of infection for the majority of drug users in Edinburgh, United Kingdom, including the three described here (15). The current sequence differences between
the study subjects therefore must have originated from a process of divergent sequence change over a period of between 9 and 10 years. By using a mean figure of 9.5 years (or 19 years of divergent sequence change), the mean synonymous pairwise distances in the gag region between individual (0.149) indicated a rate of sequence change of 0.0077 per site per year. The rate of sequence change between pairs of individuals was similar, ranging from 0.006 to 0.009 (Table 1; Fig. 3A through C). This estimate was similar to those obtained in previous studies. For example, sequence comparisons in the p1700 region of plasma RNA sequences from hemophiliacs infected from a common source indicated a mean rate of synonymous substitution in p1700 of 0.006 to 0.0072 substitutions/site/year (21).

The mean rate of nonsynonymous substitution between the study subjects was 0.0058, lower than the silent rate. The mean dS/dS ratio of 0.41 indicated a bias toward silent substitutions in this region of the gag gene, consistent with previous estimates (13, 20, 26, 34).

In this study, we also determined the sequences of the V1 and V2 hypervariable regions and flanking regions in the env region from the three study patients (positions 6560 to 6876). Between individuals, the mean pairwise synonymous distance between sequences from the flanking regions (but omitting the hypervariable regions between positions 6623 to 6679 [V1] and 6701 to 6796 [V2]) was 0.104, lower than for the p1700 region. In contrast, the rate of nonsynonymous substitution in the V1/V2 flanking region was higher (0.083), producing an overall dS/dS ratio of 0.80, similar to previous estimates for the env region (26, 48).

In this study, we used the measured rate of sequence change in the p1700 region at silent sites to estimate the time of divergence between variants infecting different tissues within an infected individual. These data should indicate when the spread of HIV into nonlymphoid tissues occurred (9).

Phylogenetic analysis of variants from different tissues. Phylogenetic analyses using sequences from the p1700 region and V1 and V2 flanking regions from a range of lymphoid and nonlymphoid tissues were carried out to determine the relatedness of variants between each tissue (e.g., lymph node, brain, and lung) (Fig. 4). Bootstrap resampling using 500 replicate trees was carried out to estimate the robustness of observed groupings.

In none of the patients was there consistent phylogenetic grouping by tissue origin. For example, p1700 sequences from lymph node samples of patient P4 were found in two distinct lineages, both of which contained a variety of sequences from other tissues (lung and spleen) (Fig. 4A). Similarly, sequences from brain samples were interspersed with those from colon, lung, and spinal cord samples. In the V1/V2 flanking regions, there was an even more marked splitting of sequences into different lineages separated by high bootstrap values (Fig. 4D). For example, sequences from the brain were found in lineages a, b, and d, of which the latter two include sequences from lymphoid tissues (lymph nodes and spleen).

Similar mixing of sequences from lymphoid and nonlymphoid tissues was observed among sequences from the other two study subjects (Fig. 4B, C, E, and F). For example, p1700 sequences from both brain and lymphoid tissue samples of patient P5 were found on lineages a and b, separated from each other by high bootstrap values (Fig. 4B). In patient P6, sequences were obtained only from brain and lymph node samples but each of the lineages contained sequences from both sources (Fig. 4C and F).

Times of divergence of HIV variants in different tissues. Pairwise synonymous distances between sequences from the p1700 region from each patient were calculated to estimate the time of divergence of variants within each tissue. The previously established rate of sequence change in the p1700 region of 0.0066 substitutions per site per year was used (20), although similar results would have been obtained if we had used the synonymous substitution rate observed in this study (mean of three study individuals, 0.0077).

Mean synonymous pairwise distance sequences within study subjects were calculated by comparing sequences from all tissues with each other, as well as comparisons restricted to variants found in particular tissues, such as brain, lymph node, and lung tissues (Table 2; Fig. 3D through F). Comparisons of variants found in all tissues produced a range of pairwise distances from 0.035 to 0.086, approximately a third of the mean interpatient silent distance. These distances implied times of divergence of 2.6 to 6.5 years (Table 2).

For all three patients, the mean distance between sequences from brain tissue was greater than the mean distance between variants in lymphoid tissue (Table 2; Fig. 5), reflecting their wide distribution in multiple lineages by phylogenetic analysis (Fig. 4). For example, the mean synonymous pairwise distances calculated for brain tissues ranged from 0.054 to 0.086 years while those for lymphoid tissues ranged from 0.035 to 0.074 years (P < 0.001). These distances translate into approximate mean divergence times of 4.1 to 6.5 years and 2.65 to 5.6 years for brain and lymphoid variants, respectively. Overall, sequences between variants found in the brain were no more similar to each other (0.080 for the three study subjects (Table

TABLE 1. Sequence comparisons between study subjects in the p1700 region

<table>
<thead>
<tr>
<th>Patient pair</th>
<th>Divergence (yr)</th>
<th>No. of pairwise comparisons</th>
<th>Mean rate of substitution</th>
<th>dS/dS ratio</th>
<th>Silent substitution rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Silent sites</td>
<td>Nonsilent sites</td>
<td></td>
</tr>
<tr>
<td>P4 and P5</td>
<td>19</td>
<td>960</td>
<td>0.152</td>
<td>0.053</td>
<td>0.35</td>
</tr>
<tr>
<td>P4 and P6</td>
<td>18</td>
<td>736</td>
<td>0.112</td>
<td>0.052</td>
<td>0.47</td>
</tr>
<tr>
<td>P5 and P6</td>
<td>19</td>
<td>689</td>
<td>0.184</td>
<td>0.070</td>
<td>0.38</td>
</tr>
<tr>
<td>All</td>
<td>18.7</td>
<td>795</td>
<td>0.149</td>
<td>0.058</td>
<td>0.39</td>
</tr>
</tbody>
</table>

* Based upon infection from a common source in 1982 (see Results).
* Silent substitution rate of sequence change (in substitutions per site per year) between study subjects.
2) than they were to those present in lymphoid tissue (mean silent pairwise distance between brain and lymph node sequences, 0.070).

Mean nonsynonymous pairwise distances were also calculated and were found to be lower than the distances at silent sites only. Nonsynonymous distances calculated for brain tissue ranged only from 0.010 to 0.042 years and were higher than those observed between variants in lymphoid tissue of the three study individuals (0.012 to 0.020; \( P < 0.001 \)). Subsequently, these values produced \( d_s/d_N \) ratios of between 0.158 and 0.49 for brain tissue only and between 0.16 and 0.46 for lymphoid tissue only, similar to those observed previously for interpatient comparisons. These ratios indicate that most of the substitutions which occurred within an individual in the p17\textsuperscript{env} region were silent.

**DISCUSSION**

**Rate of sequence change of HIV in vivo.** In this study, we used published rates as well as estimates based upon the sequences recovered from the study patients to estimate the times of divergence of variants infecting different tissues in vivo. Measurement of the rate of sequence change was possible for the study patients because it was known that all three patients were originally infected with HIV from a common source in an outbreak around 1982 or 1983, so each was in-
FIG. 4. Phylogenetic analysis of sequences obtained from different tissues of the three study subjects (P4 [A and D], P5 [B and E], and P6 [C and F]) in different subgenomic regions (p17\textsuperscript{env} region [A through C] and V1/V2 flanking regions [D through F]). Trees are shown in rooted form, with the unrelated subtype B sequence of HIV\textsubscript{AN} as an outgroup. Bootstrap values of \textgreater75\% for branches are in bold. Symbols: +, brain; ♦, spinal cord; ●, lung; ■, colon; □, lymph node; ○, spleen.

and ranged from 0.006 to 0.009 substitutions per site per year (mean, 0.0077), while the rate for the V1/V2 flanking regions was slightly lower (mean, 0.0056 substitutions per site per year).

One assumption that must be made when calculating times of divergence from sequence distances is that the rate of sequence change remains constant throughout the course of infection, and there is little direct evidence of whether this is justified. Although higher levels of virus replication clearly occur later in the course of disease, this does not necessarily imply that the rate of sequence change should be higher. The rate of sequence change is proportional to the number of replication cycles, whose length is determined by the replicative processes within the cell, unless a substantial proportion of the sampled population originates from virus that has reacted from latently infected cells, where viral replication may not have occurred for several years.

Empirically, however, the rate of sequence change at silent sites in p17\textsuperscript{env} over the first 2 years of infection in hemophiliacs (0.0066 per site per year [20]) was similar to that observed in the study patients (0.0077), in which the period of infection was 9 or 10 years, covering primary infection to death from AIDS. These figures are in turn within the range of rates of change estimated for viral isolates from several other studies (13, 26, 48).

Although the rate of nonsilent sequence change in the gag region was lower than the synonymous rate, times of divergence based on nonsilent sites were similar to times of divergence of variants in different tissues (mean time of divergence of variants within each subject, 3.1 years, compared with 3.7 years by using the published rate of silent sequence change in p17\textsuperscript{env} [20]). This is in spite of the theoretical possibility that the rate would be affected by phenotypic selection of variants with changes in the p17\textsuperscript{env} region.

By using the mean synonymous rate of substitution for p17\textsuperscript{env} of 0.0066 substitutions per site per year (20), the average time of divergence between brain and lymph node variants within an individual patient were calculated (Table 2) and a range of 3.5 to 6.5 years was obtained. In lymphoid tissue, the mean diversity of gag sequences implied an approximate population age of 2.65 to 5.6 years, while those infecting the brain were significantly more variable, suggesting an even earlier time of diversification (4.1 to 6.2 years). Despite the large potential inaccuracies in calculating times of divergence based upon sequence distances, it is clear that compared with the total duration of infection within these patients (9 or 10 years), the observed diversity within brain tissue suggests infection relatively early in the course of HIV infection and clearly preceding the onset of AIDS in two of the three study individuals (Fig. 1).

Organ-specific differences of HIV in the V3 region. Populations of HIV variants infecting different tissues in vivo are generally distinct in the V3 hyper-variable region of env (1, 23, 37) including the three study patients in the current study (8). For example, for P5, none of the V3 sequences of either the major population (15 of 17) or minor population (2 of 17) found in the brain were found among those from lung, peripheral blood mononuclear cell, and lymph node samples; these tissues were dominated by a variant with a substitution at position 28 (35 of 42). Similarly, for P6, variants in the brain were uniform and differed from lymph node variants in all but one case by 1 to 3 amino acids. The diversity of sequences in the V3 region of P4 made comparison more difficult, but again the main variants in the brain (14 of 17) were not found in peripheral blood mononuclear cell or spleen samples (n = 16).
or, with a single exception, in lymph node samples (15 sequences).

On the basis of this apparent tissue-specific distribution of variants in V3, it has been suggested that these population differences have adaptive significance and reflect different tropisms for the different infected cell types in different tissues. The involvement of V3 would be consistent with the previous observation of its role in determining the ability of HIV to replicate in different cell types in vitro (29). Within macrophagocyte isolates, an acidic amino acid or alanine was predominantly seen at position 25, while a basic or uncharged amino acid at this position was associated with nonconservative basic amino acid substitutions at positions 11, 24, and 32, correlating with T-cell tropism, consistent with the findings of other studies (5, 18, 19, 31, 41, 42, 47). Extending this work, Power et al. (37) compared cloned sequences from the brains and spleens of demented and nondemented patients and found evidence for specific amino acid substitutions at two positions in the V3 loop (histidine at position 305 and proline at position 329) that correlated with neurotropism and the clinical expression of HIV dementia. However, while other studies have also found populations infecting the brain separate from those infecting lymphoid tissues, there appear to be no conserved features of the V3 loop that correlate with neurotropism (7, 10, 21, 23, 27, 39).

Furthermore, there is no evidence for a correlation between tissue distribution with the predicted phenotype of such V3 sequences in vitro. For example, in our previous study of the three study patients and others (8), we found that each tissue was predominantly infected with variants with a predicted non-syncytium-inducing (NSI)- or macrophagcytopathic phenotype, regardless of tissue origin. In these cases, the observed amino acid differences between brain and lymphoid tissues were relatively few and probably unlikely on their own to alter the phenotype of the virus (8) (see below).

Other studies support the conclusion that the V3 region is to some extent involved in tissue tropism but that interaction with other regions in the HIV-1 genome is required for infectivity (3, 4, 22, 44). Stamatatos and Cheng-Mayer (44) have suggested that mutations altering the structure of the V3 loop can affect the conformation of gp120 and that in turn the structure of the V3 loop is influenced by the conformation of other regions in gp120. Carrillo and Ratner (3) have suggested that an interaction of the V3 loop with a small region of the C4 domain is required for infectivity of Jurkat T-cell lines, and previous studies have suggested a similar association (30, 32, 49). Therefore, although it is universally accepted that restricted variability exists in the V3 loop of HIV-1 gp120, there is no universal interpretation of this observation.

**Multiple evolutionary lineages in p1700e and V1/V2 regions.**

Given the previously observed organ-specific populations in the V3 region, we were surprised to find a different relationship between variants when sequences elsewhere in the genome were compared. In both p1700e and V1/V2 flanking regions, we observed numerous independent lineages containing sequences from nonlymphoid tissues such as the brain and lung mixed with those from lymphoid organs. Some of these groupings were confirmed by bootstrap resampling analysis (Fig. 4). Comparison of the actual V1 and V2 sequences showed a pattern of sequence variability between tissues similar to that of the flanking regions and without evidence of tissue-specific groupings; these data are the subject of a further study (17a).

There are at least three possible explanations for differences in grouping in different regions of the genome; these include (i) different rates of sequence change in different tissues, (ii) recombination, and (iii) recombination and are reviewed below.

The first hypothesis, originally proposed by Korber et al. (23), is based upon the principle that infection of nonlymphoid tissue such as the brain occurs early in the course of infection at a time when the viral population is relatively homogeneous in the V3 region. Therefore, variants infecting the brain would be initially similar to variants infecting nonlymphoid tissues. Subsequently, as disease progresses, variants found in lymphoid tissues may undergo more rapid sequence change in V3 and elsewhere in the genome associated with population replacements arising from tissue escape or antiviral treatment. For example, variants resistant to neutralization or to antivirals such as zidovudine would outgrow other variants present

<table>
<thead>
<tr>
<th>Patient(s)</th>
<th>Tissue type(s)</th>
<th>No. of sequences</th>
<th>Silent sites</th>
<th>Nonsilent sites</th>
<th>$d_{Aq}/d_A$ ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean distance</td>
<td>Mean divergence (yr)</td>
<td>$P^*$</td>
</tr>
<tr>
<td>All</td>
<td>All</td>
<td>75</td>
<td>0.049</td>
<td>3.72</td>
<td>0.031 0.390</td>
</tr>
<tr>
<td></td>
<td>Brain</td>
<td>35</td>
<td>0.080</td>
<td>6.06</td>
<td>&lt; 0.001 0.270</td>
</tr>
<tr>
<td></td>
<td>Lymph node</td>
<td>21</td>
<td>0.055</td>
<td>4.20</td>
<td>&lt; 0.001 0.185</td>
</tr>
<tr>
<td></td>
<td>Lung</td>
<td>15</td>
<td>0.049</td>
<td>3.72</td>
<td>&lt; 0.001 0.270</td>
</tr>
<tr>
<td>P4</td>
<td>All</td>
<td>20</td>
<td>0.042</td>
<td>3.18</td>
<td>0.004 0.430</td>
</tr>
<tr>
<td></td>
<td>Brain</td>
<td>7</td>
<td>0.054</td>
<td>4.10</td>
<td>0.004 0.430</td>
</tr>
<tr>
<td></td>
<td>Lymph node</td>
<td>7</td>
<td>0.035</td>
<td>2.65</td>
<td>0.004 0.430</td>
</tr>
<tr>
<td></td>
<td>Lung</td>
<td>6</td>
<td>0.070</td>
<td>5.30</td>
<td>0.058 0.430</td>
</tr>
<tr>
<td>P5</td>
<td>All</td>
<td>30</td>
<td>0.045</td>
<td>3.40</td>
<td>0.004 0.430</td>
</tr>
<tr>
<td></td>
<td>Brain</td>
<td>13</td>
<td>0.082</td>
<td>6.20</td>
<td>0.004 0.430</td>
</tr>
<tr>
<td></td>
<td>Lymph node</td>
<td>6</td>
<td>0.048</td>
<td>3.60</td>
<td>0.004 0.430</td>
</tr>
<tr>
<td></td>
<td>Lung</td>
<td>11</td>
<td>0.043</td>
<td>3.25</td>
<td>&lt; 0.001 0.185</td>
</tr>
<tr>
<td>P6</td>
<td>All</td>
<td>23</td>
<td>0.061</td>
<td>4.60</td>
<td>&lt; 0.001 0.185</td>
</tr>
<tr>
<td></td>
<td>Brain</td>
<td>15</td>
<td>0.086</td>
<td>6.50</td>
<td>&lt; 0.001 0.185</td>
</tr>
<tr>
<td></td>
<td>Lymph node</td>
<td>8</td>
<td>0.074</td>
<td>5.60</td>
<td>0.113 0.185</td>
</tr>
</tbody>
</table>

*Significance of difference between pairwise distances among brain variants compared with those in other tissues by nonparametric Kruskall-Wallis test.
within the lymphoid tissue. Rapid turnover and population replacements may be facilitated by the continuous movement of lymphocytes and other susceptible cells through lymphoid tissue. The previously estimated high rate of turnover of HIV-infected lymphocytes (14, 46) after antiviral treatment is consistent with the existence of a relatively dynamic lymphoid-cell population, whereas at least for antiviral resistance, the brain population is not (44a).

Alternatively, variants infecting lymphoid cells may be subject to more rapid changes over time associated with changes in the V3 region that determine the shift in the phenotype of HIV upon disease progression. Variants in the brain, however, may be unable to undergo such radical changes in the V3 region due to the continued strict requirements for replication in cells of the brain that are largely monocyte-derived cells, i.e., infiltrating macrophages and microglial cells (38, 45). The survival of the original infecting population in the brain and its replacement in lymphoid cells would explain the former's greater diversity in all parts of the genome other than those that determine tropism and the observed organ-specific differences in V3 populations. This hypothesis implies early entry of HIV into the brain, and although the V3 region is involved in tropism, it is not in the simple way it has been previously imagined.

It is possible to account for the organ-specific populations in V3 by other processes that do not necessarily imply early entry into the brain. For example, the organ-specific similarities in V3 sequences among variants that are not closely related in evolutionary terms could have originated from a process of strong convergent evolution, whereby the V3 sequence determines the ability of variants to grow in different cell types. Independent evidence for the existence of positive selection leading to convergence in V3 has been obtained from a study of hemophiliacs infected from a common source, who showed similarities in the pattern of sequence change in the V3 region in different individuals (20). Similarly, a longitudinal study of a single infected individual showed several independent occurrences of certain amino acid changes in the V3 loop in variants forming two evolutionarily distinct lineages (16).

It is unlikely that the V3 loop could be the sole determinant of tropism, as the differences between populations infecting brain and lymphoid tissues are often trivial and would be unlikely on their own to affect the phenotype of the virus. For example, all variants in the brain of P5 differed from those of lymphoid tissue at only one position (position 28), where a glutamate replaced an aspartate, a conservative amino acid change. Evidence for the functional equivalence of these two residues at this position can be inferred from their approximately equal distribution in isolates of the NS1 phenotype and among variants infecting a range of tissues collected at autopsy from these and other individuals (7, 10, 21, 23, 27, 37, 39).

Furthermore, if convergence were the explanation for the organ-specific grouping of V3 sequences, we might expect to observe general similarities between variants infecting specific tissues from unrelated HIV-infected individuals. However, apart from one study (37), it has generally proved impossible to demonstrate any specific conserved sequence or motif in V3 or elsewhere in env that correlates with the type of cell infected (see above). On the other hand, as noted above, it is possible that the actual V3 sequence required for replication in different cell types may depend upon interactions between V3 and other regions of env so that different V3 sequences may evolve to carry out equivalent functions in different HIV strains.

The other mechanism for different relationships in different parts of the genome is recombination, where a requirement for specific V3 sequences that confers an ability to infect different tissues may favor recombination with an already divergent preexisting population either within or without the tissue where the variants are found. Recombination occurs frequently in retroviruses, including HIV-1, and is a mechanism by which genetic variation can be increased (6). Recombination requires that multiple infection of cells occurs, and although there is evidence that this may occur in vitro (24, 40), the scarcity of HIV-infected cells in brain and other tissues seems to suggest that it may be an unlikely event in vivo. However, it is possible that recombinants are generated elsewhere, where high levels of replication occur (e.g., in lymphoid tissue), producing variants that are uniquely able to enter and replicate within the central nervous system.

Whether the similarities in V3 originated from convergence or recombination, these hypotheses suggest that the observed
diversity of variants within brain tissue could have originated by a process of multiple entry from sources outside the central nervous system. Therefore, the actual duration of infection in the brain may be substantially shorter than can be calculated by estimating its population diversity. Indeed, the grouping of variants from brain and lymphoid tissues by phylogenetic analysis of the p17Gag and V1/V2 flanking regions is evidence for a process of multiple entry. On the other hand, this hypothesis does not easily explain how populations in the brain should be consistently more diverse than those in lymphoid tissue or other presumed sources of infection in the brain (Fig. 5). The observed greater diversity of p17Gag sequences in the brain is more consistent with the first hypothesis of a lower rate of population replacement in the brain compared with that in lymphoid tissue.

In summary, the main findings of this study were the observation of an unusually diverse population of HIV variants in the brain without evidence for any closer evolutionary relationship between them than to variants infecting other tissues in the body. Although late entry of recombinant viruses is possible, it is more likely that viral entry into the brain occurs relatively early in the course of disease, based upon observations of its higher diversity in the brain than in other tissues and the existence of multiple evolutionary lineages containing sequences from the brain. These findings suggest that the loss of immune competence is not solely required for entry into nonlymphoid tissue, and the strong association between HIV-induced neuropathology and disease progression may be consequent to reactivation rather than de novo infection of the central nervous system. The finding of variants in the brain on several different evolutionary lineages challenges the hypothesis of the evolution of a uniquely neurotropic strain. It is possible that the only requirement for infection of the brain is macrophage tropism and hence the possession of a V3 loop sequence that is of low charge and shows few differences from the subtype B consensus sequence (8).

This study represents the first attempt to use evolutionary analysis of variants infecting different tissues. The finding of different interrelationships between variants in different parts of the genome, combined with uncertainty about the frequency and site of recombination in vivo and the selection pressures that could produce convergent evolution in V3, highlights the complexity in trying to understand the dynamics of HIV replication and dissemination to different tissues. However, this research at least provides a starting point for a more rigorous examination on the existence of HIV tropism in vivo.

ACKNOWLEDGMENTS

We are grateful to R. P. Brettle and C. Leen, Regional Infectious Diseases Unit, City Hospital, Edinburgh, United Kingdom, for providing clinical information on the study subjects before death. We are also indebted to staff at the Department of Neuropathology, Western General Hospital, Edinburgh, United Kingdom, and in the Hepatitis Reference Laboratory, University of Edinburgh, for the storage and provision of samples analyzed in this study. Thanks are also due to Eddie Holmes for helpful review and discussion of the manuscript. This work was funded by a Medical Research Council studentship to E.S.H. and Medical Research Council Programme grants to P.S. (PG 9209918) and J.E.B. (SPG 8925719).

REFERENCES

envelope gene is involved in macrophage tropism of a human immunodeficiency virus type 1 strain isolated from brain tissue. J. Virol. 64:6148-6153.


In Vivo Distribution and Cytopathology of Variants of Human Immunodeficiency Virus Type 1 Showing Restricted Sequence Variability in the V3 Loop

YVONNE K. DONALDSON,1 JEANNE E. BELL,2 EDDIE C. HOLMES,3 ELIZABETH S. HUGHES,1 HELEN K. BROWN,4 AND PETER SIMMONDS1,5

Department of Medical Microbiology1 and Medical Statistics Unit,2 University of Edinburgh, Medical School, Edinburgh EHS 9AG, Department of Pathology, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU,2 and Department of Zoology, University of Oxford, Oxford OX1 3PS,3 United Kingdom

Received 18 March 1994/Accepted 16 June 1994

The distribution, cell tropism, and cytopathology in vivo of human immunodeficiency virus (HIV) was investigated in postmortem tissue samples from a series of HIV-infected individuals who died either of complications associated with AIDS or for unrelated reasons while they were asymptomatic. Proviral sequences were detected at a high copy number in lymphoid tissue of both presymptomatic patients and patients with AIDS, whereas significant infection of nonlymphoid tissue such as that from brains, spinal cords, and lungs was confined to those with AIDS. V3 loop sequences from both groups showed highly restricted sequence variability and a low overall positive charge of the encoded amino acid sequence compared with those of standard laboratory isolates of HIV type 1 (HIV-1). The low charge and the restriction in sequence variability were comparable to those observed with isolates showing a non-syncytium-inducing (NSI) and macrophage-tropic phenotype in vitro. All patients were either exclusively infected (six of seven cases) or predominantly infected (one case) with variants predicted for NSI/macrophage-tropic phenotype, irrespective of the degree of disease progression. p24 antigen was detected by immunocytochemical staining of paraffin-fixed sections in the germinal centers within lymphoid tissue, although little or no antigen was found in areas of lymph node or spleen containing T lymphocytes from either presymptomatic patients or patients with AIDS. The predominant p24 antigen-expressing cells in the lungs and brains of the patients with AIDS were macrophages and microglia (in brains), frequently forming multinucleated giant cells (syncytia) even though the V3 loop sequences of these variants resembled those of NSI isolates in vitro. These studies indicate that lack of syncytium-forming ability established T-cell lines does not necessarily predict syncytium-forming ability in primary target cells in vivo. Furthermore, variants of HIV with V3 sequences characteristic of NSI/macrophage-tropic isolates form the predominant population in a range of lymphoid and nonlymphoid tissues in vivo, even in patients with AIDS.

There are several interpretations of the high degree of observed sequence diversity between published sequences of the envelope gene of human immunodeficiency virus type 1 (HIV-1) (subtype B) (54) and of the rapid turnover of HIV envelope variants within infected individuals (69, 79). It has been suggested that changes in the hypervariable domains in the env gene (V1 to V5) (52, 70, 78) may facilitate evasion of the host immune system. This conclusion is supported by the observation that V3 and possibly other regions are targets of neutralizing antibodies elicited by natural infection or upon immunization with recombinant gp120 protein (20, 26, 32, 44, 49, 60).

However, changes in the V3 loop, and more recently in the V1 and V2 hypervariable domains, have also been shown to influence the phenotype of variants of HIV-1 in vitro. In particular, substitutions of basic amino acids in the V2 and V3 regions change virus isolates from non-syncytium-inducing (NSI) isolates to syncytium-inducing (SI) isolates (11, 19, 24) and may confer a reduction in the ability of the virus to replicate in macrophages (10, 64). Apart from the association with arginine or lysine at positions 11 and 28 in V3, the SI phenotype has been also been shown to correlate with increased V3 sequence heterogeneity in this region (10, 51). V3 sequences from NSI isolates show few sequence differences from each other or from a consensus sequence of 133 North American isolates (40) that comprise predominantly subtype B variants of HIV-1 (46). In contrast, SI isolates show a broad range of substitutions, insertions, and deletions at most positions (30 of a total of 33 to 37) between the disulfide-bridged cysteine residues of the putative V3 loop structure.

In early infection, HIV variants with an NSI/macrophage-tropic (MT) phenotype predominate. Subsequently, approximately half of those individuals who progress to AIDS show a switch in virus to a more rapidly replicating (rapid/high) and cytopathic phenotype (3, 8, 18, 74); this switch in phenotype has been reported to precede an accelerated loss of CD4+ lymphocytes in the peripheral circulation and a more rapid onset of AIDS than in those individuals whose isolates retain the NSI phenotype (35). However, although most investigators have found a close association between the properties of syncytium induction and inability to replicate in macrophages in vitro and vice versa (10, 12, 19, 23, 37, 63, 64, 87), others
have found that disease progression is associated with a switch from NSI, non-MT isolates to SI/MT isolates (8).

Despite their more aggressive phenotype in vitro, the SI isolates are less transmissible by sexual contact than are NSI isolates (59, 87), but they are equally transmissible in cases of mother-to-child transmission (62). To explain these apparently contradictory findings, it has been speculated that macrophages may be the first cells infected at the mucosal barrier in the case of sexual transmission and that infection of this cell type is responsible for the systemic dissemination of the virus found upon primary infection (87). However, V3 sequences from NSI isolates were also found to be specifically selected for not only upon sexual contact but also in several cases of parenteral infection (hemophilia exposed to HIV-contaminated factor VIII (87)) in which case the initially infected cells could equally well be CD4+ lymphocytes (85).

An important limitation of many of the previous studies of sequence and phenotype change of HIV-1 is the reliance upon patient blood samples as sources of virus isolates, proviral DNA, or virion RNA as study material. However, HIV results in systemic disease and has been shown to be capable of infecting a wide variety of nonlymphoid tissue, including tissue from brain, lung, and the small and large bowel (reviewed in reference 41). It cannot be assumed that variants in blood are representative of populations infecting other cell types in the body. For example, isolates from brain and bowel tissues show in vitro properties different from those derived from peripheral blood mononuclear cells (PBMCs); the latter tend to replicate well in macrophages and to show an NS1 phenotype (25, 36, 76). In some cases, sequence differences between populations of variants in brain tissue and those in circulating lymphoid cells have been observed (16, 25, 33, 42, 43, 57, 71). The extent to which sequence changes are responsible for the differences in in vivo tropism is discussed below.

Given the evidence for changes in the in vitro phenotype of virus isolated from PBMCs at different stages of disease progression, it is clearly important to investigate whether equivalent changes occur in virus populations outside the peripheral circulation. By carrying out detailed quantitative studies of postmortem tissue, we have recently found a highly restricted distribution of HIV in the body preceding the onset of AIDS, with proviral sequences apparently confined to the lymphoid system (PBMCs, spleen, and lymph nodes). In contrast, those patients who died from complications associated with AIDS showed significant infection of cells in the central nervous system (CNS) and in lung and bowel tissues (13). These results confirm a previous study in which HIV infection was undetectable in the CNS of a large number of individuals who died while they were asymptomatic (5).

This apparent redistribution of virus upon disease progression occurs at the same stage of disease as the change from an NSI to an SI phenotype. Paradoxically, while isolates become cytopathic and often non-MT in vitro, the redistribution of HIV in vivo involves organs such as brain and lungs and other tissues in which the main targets of infection are reported to be tissue macrophages, microglia (in the CNS), and other nonlymphocyte cell types (22, 34, 47, 55, 77).

In this study, we have carried out detailed sequence comparisons of the V3 loop and flanking regions of virus variants in lymphoid and nonlymphoid tissues from a series of individuals who died while asymptomatic or as a consequence of terminal AIDS. The inferred in vitro phenotypes were compared for individuals with differing degrees of disease progression and for different tissue types in order to explore the relationship between V3 variation and tissue tropism of HIV in vivo. In many cases it was possible to directly identify the target cells and degrees of associated cytopathology in different tissues (in vivo phenotype) by immunocytochemical staining of the tissues with an anti-p24 monoclonal antibody.

MATERIALS AND METHODS

**Patient samples.** Cardiac blood and tissue from various organs were obtained within 3 days of patient death from 11 autopsied HIV-infected patients of whom three patients (patients 1 to 3) died suddenly while they were in the asymptomatic stage of HIV infection as defined by Centers for Disease Control criteria and four patients (patients 4 through 6 and patient 9) who died with AIDS-defining illness. Additional lymph node samples were obtained from four further patients who died of complications associated with AIDS (patients 10 through 13) for further sequence comparisons of sequences in the V3 and gag regions. Clinical information and laboratory investigations pertaining to patients 1 to 6 have been detailed as part of an earlier study (13). Patient 9 was a 59-year-old male with a recent history of psychomotor slowing. One week prior to death he developed a fatal atypical pneumonia, at which time a diagnosis of full-blown AIDS was made. He received no antiviral treatment, and the duration of his infection is unknown.

Samples of brain (left frontal lobe), spinal cord (midthoracic), lung, large bowel, mesenteric lymph node, and spleen tissues from these patients were dissected into 1- to 2-cm pieces and were stored at −70°C. DNA was extracted from tissues and whole blood to obtain total DNA from all nucleated cells in the circulation (peripheral blood nucleated cells [PBNCs]) as previously described (68).

**Quantitation.** Proviral DNA was quantified by using a previously described limiting dilution and nested-primer PCR approach (68). The quantitation was performed in the first instance by using primers corresponding to the pol gene and was performed subsequently with primers spanning the V3 region. The nucleotide sequences of the primers and the position of the 5′ base in the HXB2 genome (54) were pol a CATGGGTACCAAGCACCCGG, + (sense), position 4149; pol b GAGGGAAATGACAAAGTGAATA, +, 4175; pol c TCACTAGCCATGTCCTCAATT, − (antisense), 4290; pol d TCTACCTTGCTCGATGTTGCTC, −, 4380; V3 +, TA CATGGTACCAAGCACCCGG, −, 6957; V3 −, TGGCA GTCAG AAGAGAAG, +, 7009; V3 +, CATGGTACCAAGCACCCGG, −, 7331; and V3 h ATTACG TATGA AAAAAATTC CCC, −, 7381.

All reactions were performed with appropriate positive and negative controls. The number of proviral copies was estimated by assuming a Poisson distribution for each sample by −log(1−p)/d (where p = proportion of positive samples and d = dilution). Likelihood ratio tests were used to determine whether there were significant differences between different pairs. The likelihood function is proportional to \[ \left| 1 - \exp(-\lambda d) \right| ^n \times \left[ \exp(-\lambda d) \right] ^{m-n}, \] where \( \lambda = \text{number of proviral copies} \), \( m = \text{number of positive replicates} \), and \( n = \text{number of negative replicates} \). Ninety-five percent confidence intervals for each sample were determined by evaluating the likelihood function incrementally over the range of possible values for the number of proviral copies. The overall level of agreement between the pol and env primers was assessed by using the paired t test on the basis of the numbers of proviral copies on a log scale.

**Sequence analysis.** Single molecules of HIV-1 provirus were isolated by limiting dilution and amplified in a nested PCR to produce sufficient DNA to allow direct sequencing of the PCR products. Direct sequencing of amplified DNA was achieved...
either as previously described or by using a solid-phase sequencing method (30). For solid-phase sequencing, the second PCR reaction was performed in a 100-μl volume by using one biotin-labelled and one unlabelled primer (5 to 10 pmol of primer per reaction mixture), generating a PCR product with one strand having a biotin moiety at the 5’ end. PCR products were immobilized on streptavidin-coated magnetic beads (Dynal), and single strands of DNA were purified by magnetic separation and were sequenced according to the manufacturer’s protocol.

A Fisher’s exact test for 2 × n contingency tables (48) was used to compare amino acid frequencies at each position in the V3 loop (n is the number of alternative amino acids at a single position). The test is based on exact multinomial probabilities and is appropriate here because many values in the contingency tables were small. Comparisons of V3 loop charge and diversity were made between sequences obtained in this study and those of a series of isolates of HIV-1 of known biological properties. V3 loop sequences of 30 isolates with an SI phenotype and 29 NLS/MT isolates were obtained from references 10 and 19, as were corresponding sequences from MN, RF, SF2, GUN-1, HAN-2, SF33, LAI, ADA, YU2, SF162, JF-L, and SF128A isolates (51, 54).

**Evolutionary analysis of p17 sequences.** Sequence comparisons between viruses from the 11 study patients were made in the region of the gag gene encoding p17, as previously described (29). Phylogenetic relationships between single nucleotide sequences from each of the study patients and from representative sequences obtained both in Edinburgh, United Kingdom (29a), and from other widely separated geographical localities (54) were estimated by using the neighbor joining method with a bootstrap resampling program (PHYLIP programs SEQBOOT, DNADIST, NEIGHBOR, and CONSENSE) (17). Branch lengths on this tree were estimated by using the maximum-likelihood method (DNAML). Nucleotide distances between sequences were estimated by using the substitution model of Felsenstein (program DNADIST).

**Immunocytochemical staining.** Five-micrometer sections of formalin-fixed paraffin-embedded tissue from the study organs were stained with hematoxylin and eosin by standard methods. p24 antigen was detected by an avidin/biotin immunocytochemical technique (5) with a monoclonal antibody to p24 (Dupont) as first antibody at a dilution of 1/200 and diamobenzidine as the visualizing agent. Prior to antibody incubations, the sections were irradiated in an 800-W microwave oven for three cycles of 5 min each while they were immersed in 0.01 M citrate buffer (pH 6.0) (Fisons). Nucleotide sequence accession numbers. Sequences obtained in this study have been deposited into GenBank and have accession numbers L34422 to L34541.

### RESULTS

**Quantification of proviral load in different organs.** The numbers of copies of provirus per million cells in a range of lymphoid and nonlymphoid tissues from the study patients were determined by limiting-dilution PCR with primers from both pol and env regions of the genome (Table 1; results obtained with the pol primers for patients 1 to 6 have been previously reported) (13). We found high levels of HIV in lymphoid organs (spleen and lymph node) from all presymptomatic and symptomatic individuals with both sets of primers, whereas nonlymphoid organs were infected only in patients who died of complications associated with AIDS. Although low levels of provirus were found in many of the nonlymphoid tissues of the presymptomatic patients (up to 46 copies per 10⁶ cells; Table 2), such virus may have originated from PBNCs in residual blood within the organs (forming between 1 and 10% of the extracted DNA from the tissue, depending on its vascularity), as previously discussed (13). For example, the sample of colon tissue from patient 2 contained 46 copies of provirus per 10⁶ cells, 60 times lower than the frequency of infected cells in peripheral blood (2,843 copies per 10⁶ PBNCs) and attributable to the presence of peripheral blood in the

### Table 1. Quantitation of proviral sequences in postmortem tissue by limiting dilution using nested primers in the pol and V3 regions: frequency of positive results at specified dilutions of DNA

<table>
<thead>
<tr>
<th>Patient</th>
<th>Primer</th>
<th>Lymphoid tissue</th>
<th>Nonlymphoid tissue</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lymph node</td>
<td>Spleen</td>
</tr>
<tr>
<td>Presymptomatic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>E</td>
<td>9/20 (1E-3)</td>
<td>4/20 (1E-3)</td>
</tr>
<tr>
<td>2</td>
<td>E</td>
<td>10/20 (5E-4)</td>
<td>10/20 (1E-2)</td>
</tr>
</tbody>
</table>

| Symptomatic | | | | | | | | |
| 4 | E | 14/20 (5E-3) | 3/20 (1E-3) | 5/20 (1E-2) | 3/20 (1E-2) | 7/20 (1E-2) | 18/20 (1E-2) | 17/20 (1E-2) |
| 5 | E | 19/30 (5E-3) | 16/30 (5E-3) | 21/50 (5E-2) | 17/50 (2E-2) | 10/25 (1E-2) | 14/49 (2E-3) | 16/50 (6E-3) |
| 6 | E | 17/20 (1E-3) | 4/20 (1E-3) | 7/20 (1E-1) | 16/50 (3E-3) | 11/20 (1E-4) | 7/20 (1E-1) | 11/20 (1E-3) |
| 9 | E | 23/50 (5E-3) | 19/20 (1E-2) | 4/20 (5E-1) | 12/20 (2E-2) | 10/20 (1E-1) | 11/20 (1E-1) | 12/50 (8E-4) |

| | | | | | | | | |
| | | | | | | | | |

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequences</th>
<th>Accession numbers</th>
</tr>
</thead>
</table>
| E, env primer quantitation | P, pol primer quantitation |\
| Test dilution (in micrometers); 1E-3 = 1 × 10⁻³ μg | ND, not done |\
| p24² quantitation |\

"A RESTRICTED SEQUENCE VARIABILITY OF V3 IN VIVO"
5994

DONALDSON ET AL.

J. Virol.

postmortem tissue. In contrast, the majority of organs from the
r-

-3-

co

(N) O

||

o
NC
LA, GO

—1 (N

z

r\

-J ^

z

~

r\

—1
z

up

^

g
H
00

ITi

\f A ^ C O

r-o

—

h;

CM <N

y

^^^Z ^^Z
(N

ON 00

O "3"

^h

^h

lo

r-^oo^

I

o
—
O

cm

NO

on vc IT) IT,

oc oc
cm r-h

£T"
]£l
R

z—.

|
| 1^
I
I xt o
r-- r- —i cm
i

I
°

Q

j_

z

- z —z

r

Q " Q
1Z, Z
oo

-

Q
i
' ! Z
;

oo

o t, o x A f\
i
ONC--NuwC
^

cm

—i

o

in

rn

in

—

—

(n cm

co x

h

—

A

m

^ A A

'
•

•3- r-

i

no

—

oo cc o
cm cc r-

i

i

-^j-

z-v

n lol on
vz~v00
oo
•

'

>

O

I

Q ^ Q So Q

x h
co

t o h
i
o

i

co

i

T-H

cm

l

I

I

I

rH

os NC

>o
o

(N rH rH (N

rH

r- o m no
t/r
m <N Tj- OO ON O
r, 't x vc ro m

T-H

I

I

t—' -a

°
no

on cn

r- ^tI
I
rn
^

m

r-

co

on

m

fo

o lo m it, o
M Tt 0\ On
r-- no m o

1/0

^ K
on

no r-

'

in

On

o o
(N

CQ

i

-h

x) "a ^ (n co
o 't co
-3- in NO in o
hh

z

o
co

on x
cm
i
i
i 00
I —i r- m co on
NO O lO Nt (N) (N|
—'

■

u

-—-

^SCC-C'C'

ZfiZ^Z

r*t o on

i

-h
M y M y E
w^^z ^z z

—i
1

I

I

_

r-«n
.

x—s

o t-h no
^
•
•
.

.

_

_

no no co x
_
r\
00 tt
cm
o u (
^^zw^^ON 00 ^ •
o cn x r~~ cm no
i
on h- (n (n
^

'

co oo

■

NO 00

i

no cT-h CO

1

I

I

rn

Q '
1-1
'

^Z

; o!
'Z

1

•

Q
'Z
:

'ZZZ
i

no co
O r-

r-

)

O

ON

^t-

^ °
NO O i/D
00 O

1^ (N Isr
cm no T-H co
x h m h x (n x—s
CN T-H
I
I HHIO
i
i ro it
i
i no
h- o x "t x —1
i
cc, c IT) - 't ^ O A

r-l

h

VO CO
t • t

T-H

I

I

T-H

I

N h- CO
ON CM On

no on
.

I

in

O <N NO

Q T-H

no

NO NO

ON

lOhiOCOON

^

^
O

CO —1

i

ON
cr>

ON

ONr-r^ooiowocNi—)

i

oj h h h tt
1/0 co on x

m r-(J

a.

§

z
m

OJ

CX O

c

£

o

o

=
™
S

a-

PJ CU UJ 2- Uj CL

E

cqp-tua.tLi&.iao-

■c

> —

o

a.®

G.

1..S« |I Q*

E

patients with AIDS showed high levels of provirus, consistent
by HIV. Only
samples with clear evidence of infection by these criteria were
used for comparisons of sequences in the env gene (see below).
The accuracy of quantitation by limiting dilution is depen¬
dent on the number of replicates tested and the frequency of
positive reactions. We calculated the confidence intervals for
quantitation with pol and env primers by evaluating the likeli¬
hood function for each sample (see Materials and Methods)
and found that there was no statistically significant difference
between the primers for any of those tested (Table 2). When
the samples were taken together, the frequency of provirus
quantified with env and pol primers showed a ratio of 1.18
(confidence interval, 0.94 to 1.48; Fig. 1). The correlation
coefficient for quantitation with the two sets of primers was
0.96 (P = 0.0001), and there was no evidence for an over- or
underrepresentation of env or pol sequences in either lym¬
phoid or nonlymphoid tissue. These results indicate that the
V3 primers are as efficient as pol primers in amplifying HIV
and that we did not amplify only a specific subset of variants in
our subsequent sequence comparisons (see Discussion).
Sequence variation in the V3 loop. DNA extracted from
different organs or from PBNCs was diluted until only 10 to
20% of replicates gave a PCR product, thus ensuring that
nucleotide sequences were derived from single molecules of
provirus (68). The lack of peripheral infection in the asymp¬
tomatic group restricted our sequence comparisons to virus
within lymphoid tissue only (PBNCs and lymph node). For the
patients who died of complications associated with AIDS, we
were able to carry out more extensive comparisons with a
range of samples from nonlymphoid tissues, such as those from
brain and lung tissues (Table 2). None of the 322 nucleotide
sequences in the V3 loop or flanking regions contained inac¬
tivating substitutions such as stop codons or frameshifts. Only
one sequence (from lymph node tissue from patient 3) showed
G^-A hypermutation, producing a highly unusual and proba¬
bly nonfunctional provirus. Like previous researchers who
used the limiting-dilution/direct-sequencing method of se¬
quence determination (4, 68, 85), we have found no evidence
for high rates of defective genomes in vivo.
In the asymptomatic patients, a wide range of sequence
variants were found in both PBNCs and lymph nodes (Fig. 2).
In patients 1 and 3 there were statistically significant differ¬
ences in the frequency of major and minor amino acid se¬
quences present in the two types of sample. In these two
patients, the majority form, in the PBNCs, differed from that in
the lymph nodes and vice versa, although in each case such
variants were present as minority components in the other
sample.
There was a similar diversity of V3 sequences as well as of
population differences between samples from the patients who
died of complications associated with AIDS. Extreme se¬
quence diversity in all samples from patient 4 was observed,
while samples from patient 9 were restricted to only two
different sequences in the brain and to three in the lymph
nodes. As with the presymptomatic patients, the frequencies of
different variants between some organs were statistically sig¬
nificant, although a general observation was of a common set
of sequences being present at varying frequencies throughout
the body. For example, for patient 4, the main sequence in the
brain (12 of 17) and spinal cord (6 of 7) also occurred in the
colon (5 of 16), lung (7 of 15), and lymph nodes (1 of 16).
However, there are also variants that appear to be more
restricted in distribution (e.g., the colon of patient 4, the brain
of patient 5, and the colon of patient 6). This type of analysis
with actual infection of cells within the tissue

_,

ON O

^

I
ON ^

-

c-(

Nvnr(N
'

vc

-

w z m

z


provides no evidence for the existence of a shared determinant in the V3 loop that governs the distribution of HIV variants. This problem is compounded by not knowing whether the different sequences within an organ represent genuine diversity of HIV in a single cell type, or whether they result from the presence within the same organ of various proportions of different HIV-infected cell types, each bearing different proviral variants (such as CD4+ lymphocytes, dendritic cells, and macrophages in the lymph nodes).

To investigate in more detail the process of sequence diversification of the V3 loop upon disease progression, we compared the number of different V3 amino acid sequences found within samples from the presymptomatic patient group and from the group of patients with AIDS (Table 3). In both cases, frequent nonidentical sequences were found; the presymptomatic patients showed a total of 21 different sequences of 84 sampled, compared with 18 of 92 in lymphoid tissue and 23 of 146 from nonlymphoid tissue in the patients with AIDS. Similarly, the mean evolutionary distances between nucleotide sequences in the presymptomatic group (0.042) differed little from distances between sequences found in the patients with AIDS (0.038; Table 4). These comparisons also show that little difference in diversity exists between populations of HIV infecting lymphoid tissue and those replicating in nonlymphoid tissue such as that from the brain, lung, and colon (Tables 3 and 4).

Prediction of in vitro phenotype from V3 loop sequences. There exist well-defined relationships between the properties of macrophage tropism and syncytium induction on the one hand, and between the net V3 charge (11, 19) and similarity to the subtype B consensus V3 sequence (10) on the other hand. Comparison of the 54 different V3 loop sequences produced a consensus sequence that was identical in all but one position to the subtype B consensus (40). There was a striking similarity between this set of sequences and a sequence set of 29 isolates characterized in vitro as showing an MT and NSI phenotype, not only in overall consensus sequence but also in the position and nature of amino acid substitutions that did differ (Fig. 3).

In contrast, V3 loop sequences from a collection of SI sequences were highly variable and contained a number of amino acid replacements not found in the other sequences. For example, residues 7 to 9 are invariant among the postmortem and NSI/MT sequences but are highly polymorphic in the SI isolates. Other residues such as that at position 12 show differences of the consensus sequence (serine [S] in postmortem and NSI/MT sequences and arginine [R] in the SI variants). To investigate whether the postmortem sequences were significantly more similar to those of MT isolates, we carried out Fisher's exact tests (see Materials and Methods) at each amino acid position in the V3 loop. Results at each amino acid position are not independent, so statistical results have to be interpreted cautiously; values are listed only when \( P < 0.01 \). These tests showed that between postmortem and NSI/MT isolates, there were significant differences in the populations of amino acids only at positions 16 (\( P = 0.003 \)) and 22 (\( P = 10^{-5} \)) (Fig. 4). In contrast, postmortem sequences differed significantly at several positions from the SI variants, the most divergent residues being 10 (\( P < 10^{-8} \)), 12 (\( P < 10^{-8} \)), and 29 (\( P < 10^{-6} \)). This statistical evidence supports the previously observed association between the presence of basic (arginine or lysine) residues at positions 12 and 29 and the SI phenotype (19), although other types of substitutions at these sites also occur specifically in the SI variants. There are also further positions at which SI variants differ considerably from postmortem and NSI/MT variants (positions 7 to 10, 15, 28, and 36).

To investigate the relationship between in vitro phenotype and V3 sequence, we have calculated the overall V3 charge and the degree of sequence divergence from the subtype B consensus for a series of isolates with known biological properties (Fig. 5a). In agreement with a previous report which used a similar method for sequence analysis of V3 (51), NSI/MT isolates consistently showed lower charge, greater similarity to the subtype B consensus, or both, than did SI and non-MT variants. A diagonal line almost completely separates the two populations. Using this analysis to predict the phenotype of variants found in postmortem tissue of presymptomatic and terminal patients with AIDS, we found that almost all sequences were located to the left of the dividing line and that they could therefore be predicted to be of the NSI/MT phenotype (Fig. 5b and c). Indeed, there was a tendency for some of the postmortem sequences, particularly those from nonlymphoid tissue, to show a lower charge and fewer differences from the subtype B consensus than did the "typical" NSI isolate. The only sequences with overall charge and divergence approaching that of SI variants were some of those found in patients 3 and 9.

Although the sequences obtained in this study conform closely to those previously described for MT variants in vitro, it is possible that the restriction in sequence diversity in V3 was, at least in part, the result of sampling a population of individuals infected with a very limited subset of HIV-1 variants. To investigate this possibility, we carried out sequence comparisons in the p17sq for all of the 11 study patients. Phylogenetic analysis of sequences in the p17 region have been shown to provide a reliable indication of epidemiological relationships between variants within the same subtype B of HIV-1 (27, 29).
FIG. 2. Proviral V3 loop amino acid sequences from infected organs from asymptomatic patients (1 [p1] through 3) and from patients with terminal AIDS (patients 4 through 6 and patient 9). All sequences are compared with subtype B consensus sequence for V3 (40); a period indicates identity with subtype B consensus; minus indicates a gap introduced to preserve alignment with consensus sequence. n, number of sequences observed.
No direct epidemiological relationship was found between viruses infecting the 11 study subjects (Fig. 6). Although many of the subjects were infected with variants that fell within the main (heterosexual and intravenous-drug user) cluster of Edinburgh patients, there was no evidence of direct epidemiological contact between them. Furthermore, several variants were separated by those of published isolates of HIV-1 from North America and elsewhere, suggesting a distant evolutionary relationship between them. This analysis confirms the clinical impression that the patients in this study were unrelated epidemiologically to each other. However, there was no phylogenetic information within the V3 region of the env gene: sequences showed at most six amino acid changes (patient 9) from the subtype B consensus sequence. All but one V3 sequence (patient 10) is predicted to have in vitro NSI/MT phenotype according to the analysis presented in Fig. 5. Although certain V3 sequences are more divergent than others, this does not correlate with the underlying phylogenetic relationships between the corresponding p1ENV sequences within the HIV-1 B subtype.

Localization and cytopathology of HIV-1 in vivo. None of the previous investigations described in this study identify the cells infected with HIV. It is therefore not clear whether the predicted macrophage tropism and NSI phenotype of the postmortem sequences reflects the tropism and cytopathology of HIV-1 in vivo. To show this, we performed immunocytochemical detection of p24 antigen to localize the infected cells in the tissues for which we carried out the sequence analysis. For nonlymphoid tissue, we found a very good agreement between the detection of p24 antigen and virus load as determined by limiting-dilution PCR (Table 5). No p24 was found in nonlymphoid tissue of any of the presymptomatic patients, while antigen-expressing cells were frequently found in the brains, spinal cords, and lungs of several of the patients with AIDS (Tables 1 and 5). For example, patient 5 showed high levels of provirus and p24 antigen in several organs, while antigen-expressing cells in patient 6 were confined to the brain.

The lymph node architecture of the presymptomatic patients was relatively normal or showed hyperplasia (Fig. 7a), with none of the evident T-cell depletion, involution, or fibrosis which is found in the lymphoid tissue of patients with AIDS. The interpretation of p24 antigen staining in lymph nodes and spleen was complicated by the presence of extracellular virions captured on the surface of follicular dendritic cells in the B-cell areas (Fig. 7b) (15, 58). p24 antigen was confined to cellular processes of what morphologically appeared to be follicular

**TABLE 3. Comparison of sequence variability in the V3 loop upon disease progression: frequency of different sequences**

<table>
<thead>
<tr>
<th>Patient type</th>
<th>Lymphoid tissue</th>
<th>Nonlymphoid tissue</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of sequences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presymptomatic</td>
<td>84 Unique 21</td>
<td>46 Unique 23</td>
<td>130 Unique 44</td>
</tr>
<tr>
<td>Symptomatic</td>
<td>92 Unique 18</td>
<td>146 Unique 23</td>
<td>238 Unique 31</td>
</tr>
<tr>
<td>Total</td>
<td>176 Unique 36</td>
<td>146 Unique 23</td>
<td>322 Unique 47</td>
</tr>
</tbody>
</table>

**TABLE 4. Comparison of sequence variability in the V3 loop upon disease progression: mean sequence diversity**

<table>
<thead>
<tr>
<th>Patient type</th>
<th>Lymphoid tissue</th>
<th>Nonlymphoid tissue</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sequence diversity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presymptomatic</td>
<td>0.042</td>
<td>0.039</td>
<td>0.040</td>
</tr>
<tr>
<td>Symptomatic</td>
<td>0.038</td>
<td>0.039</td>
<td>0.0385</td>
</tr>
<tr>
<td>Total</td>
<td>0.040</td>
<td>0.039</td>
<td>0.040</td>
</tr>
</tbody>
</table>

*Mean evolutionary distance between nucleotide sequences.

**FIG. 3.** Comparison of sequences obtained in this study from postmortem material (54 different V3 loop sequences from a total of 322) with those of isolates of HIV-1 showing an NSI and MT phenotype (n = 29), and isolates that are SI and non-MT (n = 30). Position and frequency (in subscript) of specific substitutions are indicated below the consensus. See Materials and Methods for sources of sequences.
dendritic cells and was not found in the T-cell areas of either presymptomatic or terminal patients with AIDS (Fig. 7b). The failure to detect p24 antigen staining outside the lymphoid follicles suggests that the provirus-bearing cells in lymphoid tissue are largely transcriptionally inactive, although the p24 assay used in this study may not be able to detect low-level expression of virus proteins.

Elsewhere, p24 antigen-expressing cells in vivo were found in brain (Fig. 7c and d), spinal cord, and lung tissue (Fig. 7f and h). In brain and spinal cord tissues of patients with AIDS, p24 antigen was detected in multinucleated giant cells (Fig. 7c), in mononuclear macrophages, and frequently in morphologically normal microglial cells (Fig. 7d). In general, the presence of p24 antigen-positive cells was associated topographically with evidence of tissue damage.

In the lung tissue of patient 4, frequent p24 antigen-expressing lung macrophages were found, and, as in the brain, infected cells formed pronounced multinucleated syncytia (Fig. 7f). For this patient, the predominant V3 sequence of provirus amplified from lung tissue was identical to the major variants infecting the brain and spinal cord and to a proportion of those in colon tissue (Fig. 2). However, this sequence was not represented significantly among sequences of provirus infecting lymphoid tissue from the same patient. The other patient who by quantitative PCR (patient 5; Table 1) showed significant infection of the lungs showed a sharply contrasting tissue distribution of infection. Pathological examination of the lung revealed prominent lymphocytic infiltration into lung tissue and formation of poorly formed lymphoid follicles adjacent to bronchioles (Fig. 7g and h). p24 antigen was detected within the lymphoid follicles but not elsewhere in the lung despite the presence of some multinucleated macrophages. In this patient the population of V3 sequences in lung tissue corresponded closely to that in lymphoid tissue and was distinct from that of provirus variants infecting the brain, where infection of macrophages and microglial cells was prominent (Fig. 2).

**DISCUSSION**

Restricted sequence variability in the V3 region. An unanticipated finding in this study was the limited sequence variability in the V3 loop of HIV amplified from tissues in vivo. This restriction was evident irrespective of the degree of

![FIG. 4. Comparison of the distributions of amino acid changes at each position in the V3 loop between postmortem sequences and those of S/T-cell-tropic isolates (solid line) and those of SI/T-cell-tropic isolates (dotted line) in vivo. The probability of the distribution of changes arising by chance is indicated on the y axis on a log scale. Sequences compared correspond to those shown in Fig. 3.](image-url)
disease progression and tissue origin, whether lymphoid or nonlymphoid. Both the consensus and observed polymorphisms almost exactly matched those found in a separate analysis of isolates showing an MT/NSI phenotype in vitro, but were quite distinct from those of SI variants (Fig. 4). The infrequency with which variants with a predicted SI phenotype were detected in vivo contrasts with their frequent isolation from patients upon disease progression (3, 8, 18, 74). For example, among the four patients with AIDS in this study, only two variants were found to show sufficient positive charge and divergence from the consensus sequence to place over the dividing line that separates isolates with different properties (Fig. 5c).

We were able to rule out two possible sources of bias in our results, i.e., that the primers used for amplification of the V3 loop preferentially amplified NSI variants, and that we were studying an epidemiologically very restricted group of patients. To address the first possibility, we compared quantitation of proviral sequences by using primers specific for both V3 and the well conserved pol region. Evaluating the likelihood function to determine confidence intervals for the quantitation method used in this study, we found no significant differences between virus loads when we used the two sets of PCR primers. These data make it unlikely that we failed to amplify a significant proportion of envelope sequences and, indeed, the close concordance between the results for pol and V3 in every sample further suggests that each of the sequences analyzed was derived from a complete provirus (see below).

The second potential bias was addressed by phylogenetic analysis of p17\textsuperscript{core} sequences from the 11 patients from whom we obtained V3 loop sequences. This analysis suggested that the individuals were infected with a representative range of HIV-1 subtype B variants and that the similarities observed in the V3 loop did not result simply from infection with relatively homogeneous and possibly epidemiologically related variants.

The high frequency of variants with a predicted NSI phenotype is not inconsistent with the results of previous studies. Variants with predicted NSI phenotypes were frequently detected in several published analyses of viral sequences in vivo, in many cases from patients with advanced HIV-related disease (16, 28, 33, 37, 51, 56, 62, 67, 80, 81, 86), although their frequency relative to SI variants varies considerably between patient groups. Sequence comparisons of the V3 loop of

<table>
<thead>
<tr>
<th>Patient</th>
<th>Lymphoid tissue</th>
<th>Nonlymphoid tissue</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lymph node</td>
<td>Spleen</td>
</tr>
<tr>
<td>Asymptomatic</td>
<td>1</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>ND\textsuperscript{b}</td>
</tr>
<tr>
<td>Symptomatic</td>
<td>4</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>ND</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Graded from - (negative) to +++ (extensive staining, frequent antigen-expressing cells).

\textsuperscript{b} ND, not done.

**FIG. 6.** Phylogenetic analysis of sequences in the p17\textsuperscript{core} gene of the study patients, and comparison with V3 loop sequence diversity. (Left panel) Phylogenetic relationships represented by a rooted tree, with the HIV-1 subtype D sequence ELI as an outgroup. ○, study patients; ●, other HIV-infected Edinburgh patients. (Right panel) Majority amino acid sequence from PBNCs (P) or lymph node (L) of samples from study patients in the V3 loop; only differences from the subtype B consensus sequences (40) are shown (see legend to Fig. 2).
FIG. 7. Morphology and immunocytochemical detection of p24 antigen in lymphoid and nonlymphoid tissue. (a) Lymph node of patient 2 (asymptomatic) at low magnification (×31); hematoxylin and eosin stain. Normal germinal centers are set in the lymphocyte cortex (arrows). (b) High-magnification (×193) view of the same lymph node stained for p24 antigen, showing association of viral protein with follicular dendritic cells. (c) High-magnification view of section of frontal lobe of the brain from patient 5, stained for p24 antigen, showing prominent virus expression in giant (syncytial) cells (nuclei indicated by arrows; ×308). (d) Same case showing p24 positivity in cells of microglial morphology (arrows; ×193). (e) Section of lung from patient 4 showing frequent macrophage syncytia (arrows) in lung alveoli filled with mononuclear macrophages and edema.
sequential samples collected upon disease progression often show a change from a predicted NS1 to SI phenotype (80), although there is no evidence for a complete replacement. For example, in an HIV-infected hemophiliac monitored over a period of 6 years, V3 loop sequences were NS1 in early infection and became predominantly SI between 4 and 5 years after infection but subsequently reverted to an apparent NS1 phenotype (28, 69).

The observed restriction in sequence variability of the V3 loop can be plausibly accounted for by strong selection against sequence change in this region, although with some tolerance of certain amino acid replacements at specific sites (28, 51). The mechanism of selection for these variants remains obscure (see below), but it is significant that exactly the same restriction in sequence diversity in the V3 loop in variants associated with primary infection has been observed in vivo (38, 85, 87).

**In vitro phenotype of HIV.** Although we have not been able to confirm the phenotype of the postmortem variants experimentally, direct evidence that variants found in the CNS were indeed MT has been reported by others (9, 33). In one study, it was found that almost all isolates derived from the cerebrospinal fluid from a range of asymptomatic and symptomatic individuals were capable of efficient replication in primary macrophage culture (33); these variants showed V3 loop sequences with low charge and little divergence, if any, from the subtype B consensus.

Most of the published sequences of SI variants used for sequence comparisons were derived from isolates of HIV-1 that were often passaged extensively in cell culture prior to biological characterization. It is possible that whatever selective constraint restricts sequence diversity in vivo is absent in the conditions used for virus culture and that the virus is therefore free to drift away from the subtype B consensus sequence. It is also possible that a V3 (or V2) loop with a large positive charge confers a growth advantage in vitro, leading to the selective isolation of variants bearing such divergent sequences from a heterogeneous in vivo population. Indeed, the isolation of SI variants from patients progressing to AIDS is associated with increased virus load and therefore with a greater likelihood that such extreme variants might by chance be present in the initial PBMC culture.

Changes in the properties of isolates upon in vitro passaging are commonly observed. Repeated passaging enables HIV to adapt to efficient replication in different cell types, including permanently transformed T-cell lines. It has also been shown that in vitro culture leads to a rapid loss of sequence variability in the env gene (39) and often to the replacement of the predominant in vivo variant with a minor population (37, 39, 50). Indeed, specific outgrowth of SI variants has been found upon short-term primary lymphocyte culture of PBMCs with V3 loop sequences that could be predicted to be mainly NS1/MT (37, 47a, 61). These data are consistent with the hypothesis that the overrepresentation of SI isolates from patients with AIDS compared with their frequency in vivo is at least in part the direct result of their competitive advantage over NS1/MT variants in certain in vitro culture conditions and therefore does not necessarily reflect their prevalence in vivo.

Other lentiviruses such as HIV-2 and several of the simian immunodeficiency viruses (SIV) have envelope proteins with many structural similarities to HIV-1 gp160. In particular, it is possible to identify the homologs of the V1/V2 and V4 hypervariable regions in these different viruses. One puzzling feature has been the low degree of sequence variability in the counterpart of the HIV-1 V3 loop in HIV-2 and SIVmac (1, 7, 66). The results in this paper suggest that the designation of V3 as a hypervariable region may have been unduly influenced by the characteristics of cultured isolates of HIV-1 and may not reflect the relative homogeneity of sequences in this region in vivo. The similarities between HIV-1 and other lentiviruses may be greater than was previously imagined.

**In vivo phenotype of HIV.** The sensitive and specific detection of actively replicating virus in cells by immunocytochemical staining for p24 antigen provided an opportunity to identify the main target cells of the sequence variants identified in this study in vivo and their associated cytopathology. The only infected cells that could be identified by this technique were tissue macrophages in the lung and either macrophages or microglial cells in the CNS. This macrophage infection was associated with tissue damage and with frequent giant cell formation.

In contrast, infected cells in the lymph node were not detected despite the detection of high frequencies of proviral sequences in DNA extracted from lymphoid tissue in both asymptomatic and symptomatic patients and those with AIDS. These findings, however, are consistent with previous reports of extensive but latent infection in lymphoid tissue (14), which were based upon the finding that the number of provirus-bearing cells within a lymph node greatly exceeded the number of those in which viral RNA sequences could be detected (58). Proviral loads calculated in this study and in others using similar techniques (58) range from 10^2 to 2 × 10^6 copies per 10^6 cells. These figures are not inconsistent with the reported high frequencies of provirus-bearing cells detected by in situ PCR (14, 15), as the former figures are for total lymph node DNA, which includes nucleic acids from cells not susceptible to HIV infection (B cells, follicular dendritic cells, cells within connective tissue, etc.).

The reason why HIV infection is largely latent in lymphoid tissue remains unclear. One possibility is that the DNA detected by PCR is partially reverse-transcribed provirus produced within extracellular virions (45, 75, 84). Alternatively, the DNA may be in the form of intracellular (cytoplasmic) partial transcripts previously observed in vitro upon exposure of HIV to cells that are nonpermissive for infection (21, 82, 83). As first-strand synthesis of provirus proceeds from the 3' end of the genomic RNA, transcripts for the 3' long terminal repeat and env should be relatively more abundant than transcripts of the pol and gag genes, as has been documented in vitro (82, 84). However, in this study careful quantitation by limiting-dilution PCR with nested primers showed no difference in the relative frequencies of V3 and pol region sequences, irrespective of whether the samples were from tissue of lymphoid or nonlymphoid origin (Tables 1 and 2; Fig. 1); this suggests strongly that the sequences detected and sequenced in this study originated from complete proviruses.

**HIV tropism in vivo.** We were unable to differentiate variants in this study in terms of SI/NS1 phenotype, as all showed similar low charge and high degree of sequence conservation in the V3 loop. However, it was evident that differences existed in the populations of variants infecting fluid (×77). (f) High-magnification (×193) view showing p24 antigen-expressing syncyitia in the walls of alveoli. (g) Section of lung from patient 5 showing prominent lymphocytic infiltration (arrow) and containing carbon debris. The lung alveoli are partially filled with edema fluid (×77). (h) Distribution of p24 antigen within lung lymphoid tissue (arrow), similar to that observed in lymph nodes and spleen tissue.
different organs. For example, no patient showed equivalent distributions of variants in the CNS and in lymphoid tissue.

Although it is possible to document rapid turnover of env sequence variants with time in plasma of HIV-infected individuals (69, 79), little if anything is understood about the dynamics of sequence change outside the peripheral circulation. In particular, it is not known if separate populations of HIV develop in isolation from variants in other parts of the body (local evolution) or whether there is a process of continuous infection and spread from variants circulating in the blood (systemic evolution). Local evolution in nonlymphoid tissues might follow the widespread dissemination of HIV upon primary infection and persist at a low level (undetectable by PCR or by immunocytochemical staining) throughout the asymptomatic phase of infection because of cytotoxic T-cell activity or other immune effector mechanisms. This restriction on virus replication would become increasingly ineffective upon progression; the reactivation of virus replication in nonlymphoid tissue would form the basis for the previously observed redistribution of HIV in patients with AIDS (5, 13).

In this model, different populations would develop in different tissues, through evolutionary drift and possibly through specific adaptive changes for replication in different cell types.

However, a consistent feature of the sequence distributions in this study and in others was the dispersed nature of many of the variants. For example, for many of the study patients, the major components of populations in brain and other nonlymphoid samples were often found as minor variants in lymphoid tissue (and vice versa), suggesting repeated traffic of virus between the two in patients who were severely immunosuppressed. These findings are more consistent with the hypothesis of systemic evolution of HIV, in which virus variants disseminate freely throughout different tissues in the later stages of infection and restrict the development of local populations.

One method of virus spread documented in this paper is by lymphocyte infiltration of a tissue. In patient 5, the high proviral load detected by quantitative PCR resulted from the formation of differentiated lymphoid tissue within the lung and was associated with the presence of a virus population which was indistinguishable from that in lymph nodes and spleen tissue. In contrast, widespread dissemination of infected macrophages most plausibly accounts for the sequence similarity between variants infecting lung, colon, and brain tissue in patient 4, since the sequence identity of the V3 loop would be unlikely to have arisen by chance at several sites in the body, as would be necessary in the local model of evolution.

What remains unclear is whether variants associated with infected macrophages are functionally distinct from those found in lymphoid tissue. We are currently investigating this question through further sequence comparisons of different parts of the env gene, and through in vitro characterization of isolates derived from lymphocyte- and macrophage-infiltrated tissues.

This study provides no information on the possible origins or fate of syncytium-inducing variants of HIV that are frequently isolated upon disease progression. In this study, not only did we fail to detect variants with a predicted SI phenotype in the patients with AIDS, but the observed behavior of HIV in vivo consistently differed from that observed with cultures. Immunocytochemical staining for p24 antigen in the CNS (and lung tissue) of the patients with AIDS revealed cytopathic infection of macrophage/microglial cells by variants with a predicted NSI phenotype. These findings suggest that in vivo observations of giant cell formation are not reproduced by infection of cell lines such as MT-2. Furthermore, the ability of HIV to infect macrophages is not lost with disease progression as was previously suggested (10, 63). Indeed, the findings in this paper appear more consistent with previous findings that progression is accompanied by a change in the phenotype of isolates to SI variants that retain an ability to replicate in primary macrophage culture (8).

Independent evidence for the importance of macrophage tropism in pathogenesis has been obtained from observations of a more rapid CD4⁺ lymphocyte depletion in SCID-Hu mice infected with an MT variant than in those infected with SF-2 (53). In another animal model, the lack of disease progression, the stable CD4⁺ count, and the low circulating virus loads in chimpanzees infected with HIV-1 were attributed to a species-specific inability of HIV to infect chimpanzee macrophages (65). Directly or indirectly, the investigations described in this communication may contribute to our understanding of the mechanism by which T cells become depleted, the influence of macrophage infection, and the role of infection in the lymph nodes and spleen. These questions are essential for understanding the pathogenesis of HIV infection but are currently unresolved.

ACKNOWLEDGMENTS

We thank staff in the Department of Neuropathology and Medical Microbiology, University of Edinburgh, for technical assistance in storage and preparation of postmortem samples used in this study.

This work was funded by grants awarded by the Medical Research Council to P.S. (PG 9209918 with A. J. Leigh Brown) and to J.E.B. (SPG 8925719).

REFERENCES


47a. Mayer, A. Personal communication.


