Investigation of population diversity of human immunodeficiency virus type 1 in vivo by nucleotide sequencing and length polymorphism analysis of the V1/V2 hypervariable region of env

Elizabeth S. Hughes,1 Jeanne E. Bell2 and Peter Simmonds1

1 Department of Medical Microbiology, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, UK
2 Department of Neuropathology, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

In this study we have analysed variability in the V1 and V2 regions of human immunodeficiency virus type 1 (HIV-1) proviral sequences amplified from lymphoid tissue, brain and other non-lymphoid tissue collected at autopsy from three HIV-1-infected individuals with giant cell encephalitis. We found no evidence for any tissue-specific grouping of variants in the V1/V2 regions, in contrast to previous comparisons of sequences in the V3 region, but consistent with the existence of evolutionarily distinct lineages previously observed in these study subjects by sequence comparisons of the gag gene. Examination of inferred amino acid sequences from V1 and V2 revealed no correlations between tissue origin with overall charge, length or number of glycosylation sites. 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, we found no evidence for changes in the length of V2 with development of AIDS. Using the number of length variants in the V1 and V2 hypervariable region as a marker of the overall degree of variability within HIV populations, we found no evidence for an increase or a decrease in diversity between those with and without AIDS defining illness.

Introduction

Human immunodeficiency virus type 1 (HIV-1) replicates as a diverse, highly dynamic population in vivo (quasispecies) whose variability may contribute to its ability to survive external factors such as immune surveillance and allow it to adapt for replication in different cell types. Virus variants within such a quasispecies potentially differ in their biological properties such as cellular tropism, cytopathicity, syncytium induction (SI), replication rates and neutralization properties. A number of these differences in phenotype have been attributed to the envelope glycoprotein gp120. This envelope protein consists of five hypervariable regions (V1–V5) interspersed with four more conserved regions (C1–C4). Previous studies have implicated the V3 hypervariable domain 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, 1994; Chesebro et al., 1992). This region has also been shown to be the principal target of neutralizing antibodies produced by infection or immunization (Hwang et al., 1991).

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. V1 and V2 domains can act as targets for neutralizing antibodies, and variability in these regions may contribute to immune escape from neutralization (van Tijn et al., 1989; Moore et al., 1993; Fung et al., 1992). The V2 domain has been shown to determine efficient infection of macrophages (Chesebro et al., 1992; Toohey et al., 1995; Koito et al., 1994). Finally, the existence of functional interactions between V1/V2 hypervariable regions and other regions of gp120 has been suggested to be essential for viral infectivity and syncytium induction (Andeweg et al., 1993).

Sequences of V1 and V2 from biologically characterized populations have provided insights into the genetic diversity of HIV-1 in vivo. Studies have identified multiple evolutionarily distinct lineages among HIV-1 sequences from infected patients, and these lineages differ in their biological properties. The V1/V2 regions have been particularly useful for investigating the genetic diversity of HIV-1 in vivo, as these regions are highly variable and provide useful genetic markers for population studies.

Sequences obtained in this study have been submitted to GenBank, and have been assigned the accession numbers U79870–U79957.
isolates of HIV-1 show considerable variability in overall charge (V1, -5 to +2; V2, -3 to +4) and in length (V1, 16 to 49 amino acids; V2, 38 to 61 amino acids) (Groenink et al., 1993; Wang et al., 1995; Cornelissen et al., 1995). They also differ in the position and number of several of the potential N-linked glycosylation sites, although some are highly conserved, as are the cysteine residues that maintain the structure of the V1 and V2 loops (Leonard et al., 1990). There have been several attempts to correlate variability in V1 and V2 and biological differences such as cellular tropism and cytopathology in vitro, in an analogous way to previous investigations of V3. Groenink et al. (1993) observed a greater length and overall positive charge of the V2 domain amongst SIV, non-macrophage tropic isolates. However, subsequent investigations from this and other groups have effectively discounted a causal association between phenotype and the primary amino acid sequence of V2 (Fouchier et al., 1995; Cornelissen et al., 1995; Palmer et al., 1996; Wang et al., 1995), even though determinants of cytopathology and cellular tropism clearly are, in part, dependent on this and neighbouring regions of gp120 (Fouchier et al., 1995; Groenink et al., 1993; Andeweg et al., 1993).

Compared with the extensive analysis of V3 genetic variability in vivo, there are few corresponding studies of the V1 and V2 regions. In this study, we 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 various stages of disease, or between different infected tissues.

Methods

Study subjects. Tissues from various organs were obtained at autopsy, carried out within 3 days of death, on 42 HIV-infected individuals. Eight individuals died from other reasons and were classified as pre-symptomatic on the basis of the absence of HIV-related symptoms before death and an absence of HIV-related changes upon pathological examination. In contrast, 34 individuals died from complications associated with HIV infection, such as opportunistic infections and neoplasia (symptomatic group). Samples of brain (left frontal lobe), lymph node (mesenteric) and in some cases lung from each individual were dissected into 1–2 cm pieces and stored at -80°C for strand extension. Each template strand was subjected to 25 cycles of amplification in primary and secondary PCR reactions.

Nucleotide sequencing. 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 using one biotin labelled and one unlabelled primer (5–10 pmol per reaction) 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 protocol (Sequenase version 2.0). Following this sequencing reaction 5–6 µl of the reaction product was electrophoresed on a denaturing polyacrylamide gel (6% acrylamide, 0.38 M boric acid, 0.02 M EDTA, pH 8.3). Gels were fixed, dried and exposed overnight on BioMAX film. Sequences obtained in this study have been submitted to GenBank, and have been assigned the accession numbers U79870–U79997.

Length polymorphism analysis. 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 dNTPs was reduced from 33 µM to 8 µM and 2 µCi [α-35S]dATP was added to a reaction volume of 20 µl. Only 0.1 µl instead of 1 µl of primary PCR product was transferred to reduce the production of non-specific bands produced by carry-over of excess primary PCR product and primers. Subsequently, 1 µl of the secondary PCR product was mixed with 5 µl of stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol), heated to 90°C to denature DNA and 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 then fixed, dried and exposed overnight on BioMax film.

At the start of this study [α-35S]dATP was used for LPA. However, it has recently been shown that this radionucleotide releases H2S when heated during thermal cycling which can permeate polypropylene tubes and cause radioactive contamination. More recently, we have substituted [α-32P]dATP in all our experiments, using 0.5 µCi per reaction.

Statistical analysis. Sequence comparisons between viruses from three of the study subjects (4, 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 lengths 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 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 were considered statistically significantly different when \( P < 0.05 \) (SYSTAT version 5.0 package).

Results

Intra-sample variation in V1 and V2 hypervariable regions

We compared the 87 V1 and V2 amino acid sequences from the three HIV-infected individuals dying with AIDS (Fig. 1A–C). We found a high degree of variability in 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). 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 found in all four lineages separated by high bootstrap values (Fig. 3A), while lineages a, b and d also contained a number of specific 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). These differences are 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. 3A), 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.

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, we 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. 4). We observed a good concordance between the number and length of variants observed using LPA with sequences obtained from the three study subjects. For p4 and p5, there were differences in charge between variants from brain compared with lymph node or lung (Fig. 2A–F). The calculated overall charge of V2 of variants from brain was significantly higher than those from the lung in both patients (\( P = 0.055 \) and 0.014 respectively). 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. The significance of these differences is difficult to interpret as the distributions of values compared were derived from populations that were in some cases closely related genetically, and therefore do not constitute independent observations. There were no correlations between tissue origin with length of V1 or V2, or with number of potential N-linked glycosylation sites (data not shown). 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, we 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. 3A–C). 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). These differences are 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. 3A), 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.

V1 and V2 sequence variability and tissue tropism

We compared the overall charge, number of length variants and number of potential glycosylation sites of V1 and V2
the actual lengths of variants obtained from sequencing (Table 1), although LPA also detected minor variants not represented among the nucleotide sequences. From this initial investigation we found no consistent differences between lymph node and brain tissue. In some cases, for example p4, the same length variant was observed in both tissues. However, in patients 5
Variability of HIV V1 and V2 in vivo

Fig. 1. Proviral V1 and V2 domain amino acid sequences from the three study subjects (A) p4, (B) p5 and (C) p6. All sequences are compared with the HIVMN sequence for V1 and V2. Symbols: '.', identity with HIVMN; '-', gap introduced to preserve sequence alignment.

Fig. 2. Comparison of the overall charge of V1 and V2 sequence variants from p4 (A, D), p5 (B, E) and p6 (C, F) in the 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) are indicated by a horizontal bar.
and 6 a number of different variants were observed in both tissues.

We subsequently applied this technique to examine the diversity and overall length of V1 and V2 of variants infecting brain and lymphoid tissues from a larger study group. This comprised samples from 8 pre-symptomatic individuals at time of death and from 34 dying with AIDS (Fig. 5). Length comparisons of variants from brain were confined to those with evidence of giant cell encephalitis (GCE) since low or undetectable frequencies of proviral sequences were observed in individuals without GCE (Donaldson et al., 1994b).

We compared the number of different length variants obtained from tissues from pre-symptomatic and symptomatic individuals with CD4 counts ranging from 802 to < 1 and found no significant difference between each group in either the V1 or V2 region (Fig. 5A, B). Similarly there was no correlation between the length of variants with disease status (Fig. 5C, D). The V2 region was on average longer amongst variants from brain compared with those from lymph node tissue 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; $P < 0.001$), and pre-symptomatic patients in both V1 and V2 (mean number of length variants: V1, 4.1; V2, 3; $P = 0.039$ and < 0.001 respectively).

**Discussion**

**Lack of tissue-specific grouping by sequence variants from 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. V1 and V2 sequence variants from lymphoid and non-lymphoid tissues failed to reveal any consistent differences at specific amino acid positions (Fig. 1), in overall length, charge (Fig. 2) or overall similarity (Fig. 3).

Although the V1 and V2 sequences contain little or no phylogenetic information because of their high variability and short length, the observed lack of tissue-specific grouping was consistent with a previous study of the evolutionary analysis and tissue distribution of the p17 gag region from the same three individuals (Hughes et al., 1997). In this previous study we found multiple evolutionary lineages in both of these regions of the HIV-1 genome. Using the p17 gag region an estimation of the time of diversification of in vivo variants from brain tissue was calculated to be 4.1 to 6.2 years, suggesting infection of brain tissue may occur as an early event in disease and preceding the onset of AIDS.

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. 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
Variability of HIV V1 and V2 in vivo

Fig. 4. LPA of HIV variants obtained from lymph node and brain tissue of p4, p5 and p6 for (A) V1 and (B) V2. The comparative lengths of sequences are indicated as the number of amino acids.

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 these biological characteristics 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 for a difference in length of V2 between pre-symptomatic and AIDS study subjects, despite the greater expected frequency of isolation of SI variants from the latter group. However, in a previous analysis of the V3 region from the three patients in this study and others the majority of sequences were found to have a predicted NSI/macrophage tropic phenotype (Donaldson et al., 1994).

These findings suggest that V1 and V2 may contribute to tissue tropism in a 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 to facilitate infection by HIV-1 (Andeweg et al., 1993; Koito et al., 1994; Groenink et al., 1993). It seems unlikely that genomic regions which have been shown to influence virus 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 residues that determine these properties.

The lack of organ-specific groupings contrasts with previous comparisons in the V3 region in both our own and other investigations (Reddy et al., 1996; Donaldson et al., 1994a) 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. If infection of non-lymphoid tissue such as brain occurs early on in disease progression (Korber et al., 1994) the virus population would be relatively homogeneous in V3 and hence non-lymphoid and lymphoid tissues would be initially expected to harbour similar virus variants. However, as disease progresses, lymphoid variants may undergo more rapid sequence turnover and population replacement resulting from immune pressure and/or antiviral drug treatments. Variants harboured within tissues such as brain may either be less exposed to such selection pressures or alternatively the requirements for infection of cells of the brain, which are largely monocyte derived (Vazeux et al., 1987; Price et al., 1988), may prevent the degree of sequence change in V3 tolerated by infection of lymphoid tissue. Indeed, a recent study (Wong et al., 1997) has provided evidence to suggest an extremely slow turnover of variants infecting brain tissue (see below). The lack of segregation of V1 and V2 sequences observed in this study suggests either that these two regions of the HIV-1 genome do not influence the tropism of HIV-1 to the same extent as V3 or that the former regions may influence tropism in a way that is not discernible from their primary sequence.

Comparison of the rates of synonymous and non-synonymous substitution rates in the V1 and V2 regions of env produced dS/dN 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 (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 (Li et al., 1988; Kasper et al., 1995) for p17 gag, a region characterized 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 have produced ratios of 1.5 to 2.9 (Lukashov et al., 1995; Simmonds et al., 1990a). 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 neutralizing antibody recognizing 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., 1990a). This hypothesis appears even more likely for V1 and V2, given the existence of neutralizing epitopes in this region (Moore et al., 1993; Fung et al., 1992; van Tijn et al., 1989). Van Tijn et al. reported the detection of antibodies to a V1 peptide, corresponding to residues 19 to 34 in our numbering system, in eight 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 two or three glycosylation sites which may affect envelope conformation. Indeed, Gram et al. (1994) 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.

Both McKeating et al. (1993) and Warrier et al. (1994) have described epitopes in the V2 region encompassing residues 60 to 70 in our numbering system. This region is relatively well

| Subject | Organ | 27 | 30 | 31 | 32 | 35 | 36 | 38 | 40 | 41 | 42 | No.| 27 | 30 | 31 | 32 | 35 | 36 | 38 | 40 | 41 | 42 | No. |
|---------|-------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| p4      | LN    |    |    |    |    |    |    |    |    |    |    | 1  |    |    |    |    |    |    | 12 |    |    |    | 1  |
|         | BR    |    |    |    |    |    |    |    |    |    |    | 1  |    |    |    |    |    |    | 10 |    |    |    | 1  |
| p5      | LN    | +  | +  |    | +  |    | +  |    |    |    |    | 4  |    |    |    |    |    |    | 5  |    |    |    | 1  |
|         | BR    | +  | +  |    | +  |    | +  |    |    |    |    | 5  | 1  | 1  | 3  | 1  | 2  | 2  | 9  | 1  | 1  | 1  | 6  |
| p6      | LN    | +  | +  |    | +  | +  |    |    |    |    | 5  | 2  | 5  |    |    |    |    | 2  | 7  |    | 1  | 1  | 1  | 4  |
|         | BR    | +  | +  |    | +  | +  |    |    |    |    | 4  | 3  |    |    |    |    |    | 1  | 1  | 1  | 4  |    |    |    |

V2 Region

| Subject | Organ | 38 | 40 | 41 | 43 | 44 | 47 | 48 | No. | 38 | 40 | 41 | 43 | 44 | 47 | 48 | No. |
|---------|-------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| p4      | LN    | +  |    |    |    |    |    |    |    | 2  | 12 |    |    |    |    |    |    | 1  |    |    |    | 1  |
|         | BR    | +  |    |    |    |    |    |    |    | 2  | 10 |    |    |    |    |    |    | 1  |    |    |    | 1  |
| p5      | LN    | +  | +  |    | +  | +  |    |    |    | 5  |    |    |    |    |    |    |    | 1  | 6  |    | 2  |    |    |    | 2  |
|         | BR    | +  | +  |    | +  | +  |    |    |    | 4  |    |    |    |    |    |    |    | 9  | 1  | 1  | 1  | 3  |    |    |    |
| p6      | LN    | +  |    |    |    |    |    |    |    | 2  | 1  | 10 |    |    |    |    |    | 2  | 1  | 10 |    | 2  |    |    |    |
|         | BR    | +  |    |    |    |    |    |    |    | 2  | 1  | 10 |    |    |    |    |    | 2  | 1  | 10 |    | 2  |    |    |    |

Table 1. Comparison of lengths of V1 and V2 obtained by LPA compared with direct sequencing

V1 Region

| Subject | Organ | 27 | 30 | 31 | 32 | 35 | 36 | 38 | 40 | 41 | 42 | No. |
|---------|-------|----|----|----|----|----|----|----|----|----|----|----|----|
| p4      | LN    |    |    |    |    |    |    |    |    |    |    | 1  |
|         | BR    |    |    |    |    |    |    |    |    |    |    | 1  |
| p5      | LN    | +  |    |    |    |    |    |    |    |    |    | 4  |
|         | BR    | +  |    |    |    |    |    |    |    |    |    | 5  |
| p6      | LN    | +  |    |    |    |    |    |    |    |    |    | 5  |
|         | BR    | +  |    |    |    |    |    |    |    |    |    | 4  |

V2 Region

| Subject | Organ | 38 | 40 | 41 | 43 | 44 | 47 | 48 | No. | 38 | 40 | 41 | 43 | 44 | 47 | 48 | No. |
|---------|-------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| p4      | LN    | +  |    |    |    |    |    |    |    | 2  | 12 |    |    |    |    |    |    | 1  |    |    |    | 1  |
|         | BR    | +  |    |    |    |    |    |    |    | 2  | 10 |    |    |    |    |    |    | 1  |    |    |    | 1  |
| p5      | LN    | +  | +  |    | +  | +  |    |    |    | 5  |    |    |    |    |    |    |    | 1  | 6  |    | 2  |    |    |    | 2  |
|         | BR    | +  | +  |    | +  | +  |    |    |    | 4  |    |    |    |    |    |    |    | 9  | 1  | 1  | 1  | 3  |    |    |    |
| p6      | LN    | +  |    |    |    |    |    |    |    | 2  | 1  | 10 |    |    |    |    |    | 2  | 1  | 10 |    | 2  |    |    |    |
|         | BR    | +  |    |    |    |    |    |    |    | 2  | 1  | 10 |    |    |    |    |    | 2  | 1  | 10 |    | 2  |    |    |    |

* The bold cross ( ) indicates more prominent bands on LPA (see Fig. 4).
† Total number of different length variants detected by LPA or by direct sequencing.
‡ Figures refer to the number of sequences obtained for each length.
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., suggesting this region may be of importance during infection. McKeating et al. also described a number of conformational epitopes located in the carboxy-terminal of the V2 loop. The carboxy-terminal sequence from the three patients studied here showed 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, variants infecting patient 5 were highly glycosylated within this region (one to four glycosylation 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.

As well as amino acid sequence variation, immune escape from neutralizing antibody may occur through alterations in conformational epitopes, including alterations in the position and number of glycosylation sites. There was substantial variability in V1 that may form sites for O-linked addition. Using enzymatic deglycosylation Bernstein et al. (1994) found evidence for the modification of gp120 by O-linked carbohydrates in addition to N-linked carbohydrates. Although the sites for O-linked addition were not experimentally determined, 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). In this study, frequent serine/threonine-rich insertions were observed in the V1 region which, although differing in sequence to those found in variants of SIV, may produce major alterations in the pattern of glycosylation and hence antigenicity of this region. Furthermore, syncytium formation by HIV-1 has also been shown to be blocked by antibodies to O-linked carbohydrate structures (Hansen et al., 1991).

**Does the diversity of V1 and V2 sequence variants correlate with disease progression?**

We 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. 4).

LPA provides only a partial description of the variability within a sample. While each length variant represents a dif-
different amino acid sequence, often quite diverse sequences may have the same overall length, as in V2 for p4 (Fig. 4 B). 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. In this respect, the method provides a more complete analysis of diversity than sequencing-based approaches (Table 1; Fig. 4). 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 cannot be detected by LPA, related techniques such as the heteroduplex mobility assays that have also been used to investigate in vivo variability do not differentiate between the majority of silent substitutions from those that change amino acid sequences. Variability in these assays can therefore not be precisely equated to changes that influence virus phenotype.

A number of previous studies have related sequence change and diversification of HIV-1 with disease progression. Diversification itself may lead to increasing antigenic variation and an inability of the immune system to respond (Nowak et al., 1991). Nowak et al. proposed the ‘antigenic diversity threshold’ model to explain virus diversity and disease progression. We observed no trend for either an increase or decrease in the number of length variants in V1 or V2 from lymphoid tissue with declining CD4 count at time of death, or between pre-symptomatic and symptomatic study subjects (no comparison could be made for brain tissue between the two study groups). We were therefore unable to experimentally confirm the evidence based upon sequence comparisons of much more restricted numbers of individuals for a decline in population diversity upon disease progression in regions of env that includes V1 and V2 (Lukashov et al., 1995; Ganeshan et al., 1997; Wolinsky et al., 1996). One of the difficulties of interpreting our 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 has been obtained by sequence 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. Recently, evidence for extremely slow turnover of variants in the brain was obtained by sequence comparisons of the pol gene of variants infecting brain and lymphoid tissue of study subjects with AIDS at death (Wong et al., 1997). Despite frequently prolonged antiviral treatment before 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. 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 virus 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 indeterminable by analysis of their primary sequences.

We are grateful to Dr R. P. Brettle and C. Leen, Regional Infectious Diseases Unit, City Hospital, Edinburgh 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 and in the Hepatitis Reference Laboratory, University of Edinburgh for the storage and provision of samples analysed in the study. The work was funded by a Medical Research Council studentship to E.S.H. and Medical Research Council Programme grant to P.S. (PG 9209918, with A. J. Leigh Brown) and J.E.B. (SPG 8925719).

References


Donaldson, Y. K., Bell, J. E., Holmes, E. C., Hughes, E. S., Brown, H. K. & Simmonds, P. (1994a). In vivo distribution and cytopathology of
variants of human immunodeficiency virus type 1 showing restricted sequence variability in the V3 loop. *Journal of Virology* 68, 5991–6005.


E. S. Hughes, J. E. Bell and P. Simmonds


Received 21 April 1997; Accepted 23 June 1997