JOURNAL OF VIROLOGY, Sept. 1993, p. 5623–5634
0022-538X/93/095623-12$02.00/0

Vol. 67, No. 9

V3 Loop of the Human Immunodeficiency Virus Type 1 Env Protein: Interpreting Sequence Variability

LYNN MILICH,[1] BARRY MARGOLIN,[2,3] AND RONALD SWANSTROM[1,3]*

Department of Biochemistry and Biophysics,[1] Department of Biostatistics,[2] and Lineberger Comprehensive Cancer Center,[3] University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7295

Received 6 January 1993/Accepted 10 June 1993

Two different states of human immunodeficiency virus type 1 are apparent in the asymptomatic and late stages of infection. Important determinants associated with these two states have been found within the V3 loop of the viral Env protein. In this study, two large data sets of published V3 sequences were analyzed to identify patterns of sequence variability that would correspond to these two states of the virus. We were especially interested in the pattern of basic amino acid substitutions, since the presence of basic amino acids in V3 has been shown to change virus tropism in cell culture. Four features of the sequence heterogeneity in V3 were observed: (i) approximately 70% of all nonconservative basic substitutions occur at four positions in V3, and V3 sequences with a basic substitution in at least one of these four positions contain approximately 95% of all nonconservative basic substitutions; (ii) substitution patterns within V3 are influenced by the identity of the amino acid at position 25; (iii) sequence polymorphisms account for a significant fraction of uncharged amino acid substitutions at several positions in V3, and sequence heterogeneity other than these polymorphisms is most significant at two positions near the tip of V3; and (iv) sequence heterogeneity in V3 (in addition to the basic amino acid substitutions) is approximately twofold greater in V3 sequences that contain basic amino acid substitutions. By using this sequence analysis, we were able to identify distinct groups of V3 sequences in infected patients that appear to correspond to these two virus states. The identification of these discrete sequence patterns in vivo demonstrates how the V3 sequence can be used as a genetic marker for studying the two states of human immunodeficiency virus type 1.

The viral glycoproteins displayed on the surface of enveloped viruses are responsible for interacting with the receptor on the target cell and are also exposed to humoral selection by circulating antibodies. Selection by antibodies can result in the appearance of sequence heterogeneity within discrete regions of the glycoprotein sequence, as has been seen with the influenza virus HA protein, for which the biological consequence of this heterogeneity is antigenic drift (66, 77, 80). Genomic diversity among independent human immunodeficiency virus type 1 (HIV-1) isolates (reference 68 and references therein), to a lesser degree among sequential isolates from the same patient (28, 30, 65), and even within a single patient isolate (54) is a well-characterized feature of HIV-1. Although this sequence heterogeneity is distributed throughout the genome, most of the heterogeneity is located in the *env* gene (13, 27, 50). Comparison of predicted amino acid sequences from several different isolates revealed that amino acid heterogeneity is clustered in five variable regions (V1 through V5) of the surface glycoprotein gp120 (43) (Fig. 1). When *env* gene sequences from sequential isolates were examined, amino acid changes in gp120 were found to occur predominantly in the same highly variable regions, suggesting that sequence variability may play a significant role in generating mutants capable of escaping neutralization (43).

During disease progression, a more virulent strain of HIV-1 emerges, suggesting that HIV-1 exists in two different states early and late in infection (2, 6, 19, 59, 74, 75). These states have been measured by the replicative capacity of the isolate, defined as slow-low and rapid-high; by cytopathic abilities, usually defined as non-syncytium inducing (NSI) and syncytium inducing (SI); or by alterations in host range, including virus growth in macrophages (macrophage tropic) versus adaptation to growth in transformed T-cell lines (T-cell-line tropic) (2, 6, 18, 19, 59, 60, 61). Although these three designations refer to different properties of HIV-1, it is likely that rapid-high, T-cell-line tropic, and SI all define the same state, which is distinct from a less-virulent state characterized as slow-low, macrophage tropic, and NSI.

The biological significance of the sequence heterogeneity in V3, the third variable region of gp120, has come under study (reviewed in reference 24). Although this region is only 35 amino acids long, considerable sequence variability exists within it when sequences from different isolates are compared (37). In spite of this variability, V3 contains determinants that mediate virus interactions with $CD4^+$ cells, the primary targets of HIV-1 infection, and with the host immune system. Antisera to recombinant Env proteins or to synthetic peptides that include amino acids within V3 are able to block syncytium formation induced by a virus with the homologous V3 sequence (25, 32, 48, 53). V3 also serves as a major target for neutralizing antibody (29, 32, 33, 41, 48, 53). Antibodies to the V3 region do not prevent binding of gp120 to the CD4 receptor but do prevent virus uptake (67). There have been additional reports which show that V3 contains epitopes that elicit both cytotoxic T-lymphocyte responses (10, 69, 71, 72) and helper T-cell responses (11, 49, 70). Finally, molecular recombinants have been used to show that determinants of virus tropism lie within the V3 region (3, 8, 9, 31, 46, 62, 78). Consequently, the heterogeneity characteristic of V3 may represent in part the sum of selective pressures associated with these different properties.

Initial attempts have been made to identify specific patterns of sequence variability in V3 associated with virus tropism and the cytopathic abilities of both laboratory-

* Corresponding author.

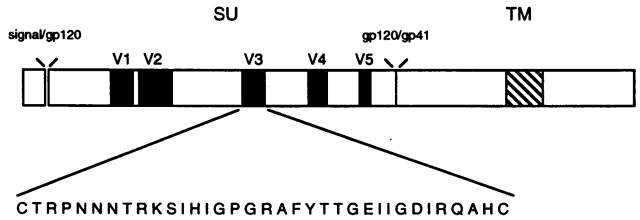

FIG. 1. Variable domains of the HIV-1 Env protein. The five variable regions (V1 through V5) in the gp120 (SU) portion of the *env* gene are indicated by black boxes. Shown below is the V3 consensus sequence (37). The shaded box represents the membrane-spanning domain within the gp41 (TM) portion of the Env protein.

adapted and primary isolates of HIV-1. Several studies have shown that coupled changes flanking the GPGR sequence, which is present in most V3 sequences, are sufficient to change virus tropism to T-cell-line tropic (9, 15). Mutant viruses have been used to demonstrate that a minimum of three amino acid substitutions in V3 can confer macrophage tropism and alter T-cell-line tropism (63). de Jong et al. (15) provided evidence that HIV isolates with an SI, fast-replicating phenotype have a higher net charge in V3 because of the addition of basic amino acids. Comparison of V3 sequences from sets of biological clones isolated sequentially from 12 patients showed a correlation between a basic charge at two positions (positions 11 and 25) and the transition from an NSI to an SI phenotype (21). The importance of these two positions was confirmed by mutational analysis (14).

We analyzed more than 250 published V3 sequences to identify sequence patterns that distinguish between the two states of HIV-1. Our starting point for the analysis of sequence patterns was two large data sets of V3 sequences, which included examples of macrophage-tropic and T-cell-line-tropic viruses (37, 44). In the second part of this study, we used the sequence patterns identified from the published sequences to analyze the virus population in infected patients.

MATERIALS AND METHODS

Source of V3 sequences. The 175 V3 sequences used in the initial analysis were selected from the data set published by LaRosa et al. (37) with corrections (38, 39) and are described in the legend to Fig. 2. An additional 83 V3 sequences obtained predominantly from the HIV-1 data base (44) were used in the analysis of V3 sequence variability and are available upon request. These sequences were determined for isolates from 64 individuals and represent both in vivo-derived sequences and sequences from cultured virus. In almost all cases, no more than two sequences for isolates from a single patient were included, representing the predominant macrophage-tropic-like sequence and a distinct sequence variant when present. The viruses were isolated predominantly from the blood, although some from the brain and the lung are included. When HIV-1 was transmitted between individuals, the isolate(s) from only one patient was used. All of the isolates in the data set except SF1703 (7) represent isolates from North America or Europe. A subset of this data set which includes 10 pairs of in vivo-derived sequences is further described in Fig. 3B.

Viral DNA isolated from peripheral blood mononuclear cell (PBMC) DNA of two HIV-1-infected individuals was used as a source for the patient V3 sequences described in Fig. 6. Patients A and B were enrolled in the AIDS Clinical Trials Group at the University of North Carolina Hospitals, Chapel Hill, N.C. Both patients had $CD4^+$ cell counts of less than 100/mm^3 and had received antiviral therapy.

PCR amplification and cloning. PBMCs were isolated from patient blood by centrifugation on Ficoll-Hypaque as previously described (12). Approximately 10^7 PBMCs were resuspended in 0.28 ml of 1× STE (0.15 M NaCl, 10 mM Tris [pH 8.0], 1 mM EDTA). Lysis buffer (500 μg of proteinase K per ml, 0.5% sodium dodecyl sulfate [SDS], 10 mM Tris [pH 7.5], 30 mM NaCl, 20 mM EDTA) was added to a final volume of 0.5 ml, and the cells were incubated for 2 h at 50°C. Samples were then phenol extracted and ethanol precipitated.

Nested polymerase chain reaction (PCR) was used to amplify the 3-kb *env* gene from the uncultured PBMC DNA. The first 30 cycles were done with outer primers 5411/5428 (sequence derived from HXB2R; 5′-AGC ATC CAG GAA GTC AGC-3′) and 8546/8565 (5′-GTA CCT GAG GTG TGA CTG GA-3′) (51). Either 0.5 or 1 μg of the patient DNA sample was used in a 100-μl reaction that included 0.5 μM each of the primers, 800 μM total deoxynucleotide triphosphates, 2.0 mM $MgCl_2$, 10 mM Tris-HCl [pH 8.3], 50 mM KCl, 2.5 U of Amplitaq [Perkin Elmer Cetus]). All 30 cycles consisted of a 30-s denaturing step at 94°C, a 30-s annealing step at 60°C, and a 2-min extension step at 72°C except for the first cycle, in which reactions were both denatured and extended for 5 min. After 30 cycles, the reactions were extended for an additional 7 min. Three successive 10-fold dilutions were made from the amplified DNA, and an aliquot of 8 μl of each dilution was placed into a new 100-μl reaction mixture for a second 30 cycles, during which a region of DNA was amplified from the previously amplified DNA. The second round of PCR was done with nested primers 5503/5537, containing an additional *Xba*I site (5′-AGC TTC GAG ATC TAG ATT AGG CAT CTC CTA TGG CAG GAA GAA GCG GAG ACA-3′), and 8421/8463, containing an *Xho*I site and an additional *Mlu*I site (5′-TAA GTA CTA TAC GCG TCA TGT TTT TCC AGG TCT CGA GAT GCT GCT CCC ACC CCA TCT GC-3′). The parameters for the second round of amplification were the same as for cycles 2 through 30 except that the denaturing step in cycle 31 was 5 min and the $MgCl_2$ concentration was 1.5 mM.

Only PCR reactions that were not overamplified were used to avoid cloning *env* heteroduplexes that may have formed by reannealing at high template concentrations. The reaction mixtures were purified by binding to glass beads by the manufacturer's protocol (Geneclean II; Bio 101 Inc.), and the amplified DNA was digested with the appropriate restriction enzymes. DNA fragments were ligated into the phagemid pIBI31 (International Biotechnologies, Inc.), and bacterial transformation was done by electroporation (16).

Recombinant clones containing the 3-kb insert were screened on indicator plates containing 100 μg of ampicillin per ml, 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and 40 μg of 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal) per ml. Between 10 and 20% of the white colonies (lacking *lacZ*) were transferred to nitrocellulose (Schleicher & Schuell BA85) and screened for HIV-1 *env* DNA by hybridization with an oligonucleotide probe as described before (55, 84) except that the filters were hybridized at 80°C in 10 ml of prehybridization solution containing 2.3 pmol of a ^{32}P-5′-end-labeled oligonucleotide (56) and cooled slowly to room temperature overnight. Plasmid clones identified in the hybridization screen were then screened for the presence of full-length *env* inserts.

***Taq* error frequency.** An error rate for *Taq* DNA polymerase was determined after 60 cycles of amplification. Roughly 100 copies of pIBI20 containing the HIV-1 *env* gene were mixed with 1 μg of uninfected high-molecular-weight human cell DNA, and the *env* gene was amplified for 60 cycles under the conditions described above. The amplified fragment was cloned and sequenced. V3 sequences were obtained for 22 clones totalling 1,980 bp. One C-to-T transition was identified, for an estimated error frequency of approximately 0.05% after 60 cycles of amplification.

DNA sequencing and sequence alignments. Single-stranded DNA for sequencing was generated after rescue of the *env*-containing phagemid with helper phage M13K07 (76). The sequences of the clones were determined by dideoxy sequencing with Sequenase Version 2.0 according to the manufacturer's protocol (United States Biochemical). The sequencing primer used for V3 was 5'-AAC CAT AAT AGT ACA-3' (sequence derived from HXB2; primer 6616) (51). The sequence analysis programs SeqEd, Translate, and PileUp (23) were used to translate and align patient V3 sequences.

Statistical analysis. All *P* values were generated by using the Fisher exact test as the significance test criterion (20).

Nucleotide sequence accession numbers. Patient sequences were deposited in GenBank under accession numbers L21769 to L21836, L21980, and L21981.

RESULTS

We examined two large data sets of published V3 sequences to identify patterns of sequence variability that can be used to distinguish between the two different states of HIV-1. Since the work of de Jong et al. and others (9, 14, 15, 21, 63) has shown that the addition of a basic charge (by substitution) in V3 is associated with the more virulent state of HIV-1, we were especially interested in determining the distribution of basic amino acid substitutions within the V3 region. In an initial analysis, we examined the pattern of basic amino acid substitutions within 175 distinct V3 sequences published previously by LaRosa et al. (37). We chose the LaRosa data set as our starting point for several reasons. First, the consensus sequence of this data set is the same sequence as that of the HIV-1 isolates BaL1 and JR-FL (44, 46), both of which are known to confer a macrophage-tropic phenotype (22, 31, 46). Second, an example of this sequence is present in the list of V3 sequences, as are examples of variants that are known to grow well in T-cell lines. We assume that this extensive list contains numerous examples of each type of virus.

Patterns of basic amino acid substitutions within V3. Analysis of the V3 sequences from the LaRosa data set show that basic amino acids appear frequently, both within the consensus sequence (at five positions) and as substitutions from the consensus sequence (Fig. 2A). Some positions tolerated conservative substitutions (9, 10, and 18), while two positions with arginine as the predominant amino acid had lysine substitutions only rarely (positions 3 and 31). Nonconservative substitutions to basic amino acids were also discriminating. Basic substitutions were most frequently seen at positions 11, 13, 19, 23, 24, 25, and 32, with substitutions at the four most common positions, 11, 24, 25, and 32, representing approximately 67% of all nonconservative basic substitutions. Basic substitutions at positions 13, 19, and 23 account for about 19% of all nonconservative basic substitutions. Nonconservative basic amino acid changes were also distributed asymmetrically at these seven positions. Four positions, 19, 24, 25, and 32, tolerated either arginine or lysine, while at the other three positions (11, 13, and 23), arginine predominated. At position 32, basic amino acid substitutions were the only substitutions tolerated.

Analysis of an additional 83 V3 sequences, including sequences from the HIV-1 data base (see Materials and Methods), revealed similar basic amino acid substitution patterns (data not shown). When V3 sequences from the two data sets were combined, we found that 70% of all nonconservative basic substitutions occurred at the four noted positions, 11, 24, 25, and 32. In addition, the subset of V3 sequences (approximately 42%) that contained a basic amino acid at one of these four positions contained approximately 95% of all of the nonconservative basic substitutions present within the total list of V3 sequences. Thus, basic substitutions at four positions (11, 24, 25, and 32) can be used as a reliable marker to identify virtually all sequence variants with nonconservative basic substitutions. In addition, basic substitutions in two of these four positions (11 and 25) have been directly implicated in virus tropism (14, 21, 63, 79).

In analyzing the V3 sequences from the LaRosa data set

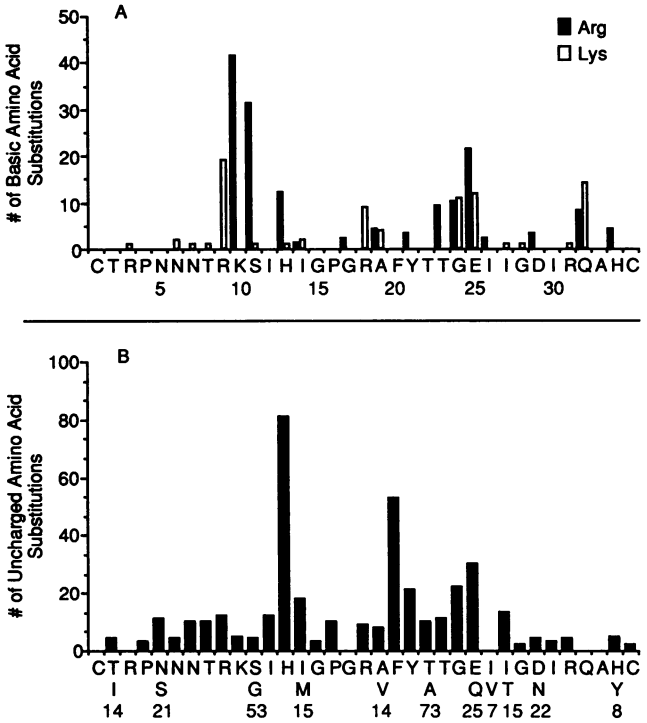

FIG. 2. Distribution of basic and uncharged amino acid substitutions within the LaRosa V3 sequence data set (37). Included in the data set are 175 unique V3 sequences published by LaRosa et al. (37). Identical sequences were used only once. Not included were 11 V3 sequences with either a QR or RG insert between Ile-14 and Gly-15 of the consensus sequence (LaRosa V3 sequences 166 to 175 and 225). Also not included were nine V3 sequences with other unusual insertions or deletions (LaRosa V3 sequences 64 to 66, 178 to 180, 212, 223, and 244). Seventeen sequences which contain a single amino acid deletion at position 18, 22, 24, or 27 were included in the data set (LaRosa V3 sequences 39, 63, 101, 121, 122, 133, 135, 144, 163, 164, 208 to 211, 222, 232, and 245). (A) Number of arginine and lysine substitutions. (B) Number of uncharged amino acid substitutions and deletions not including polymorphisms. Specific polymorphisms are shown below the consensus sequence in panel B. Numbers indicate how many times each polymorphism was found among the 175 sequences.

that had nonconservative basic amino acid substitutions, we also noted that most of them also had a sequence change at position 25, away from the consensus glutamic acid or its conservative substitute, aspartic acid. This correlation was also apparent if the sequences were first grouped as being substituted at position 25. Of the 88 sequences in the list that contained a nonacidic amino acid at position 25, 69 of them had a basic amino acid at position 11, 13, 19, 23, 24, or 32. By contrast, only 10 V3 sequences on the list had a basic amino acid at one of these six positions and an acidic amino acid at position 25. Thus, there was a significant correlation for a change away from an acidic amino acid at position 25 and the appearance of basic amino acids at the other six positions ($P < 10^{-4}$).

Patterns of acidic and uncharged amino acid substitutions within V3. In addition to identifying basic amino acid substitution patterns that may be associated with virus state, we also wanted to address the broader question of sequence variability within V3. To do this, we examined the distribution of acidic and uncharged amino acids in the 175 V3 sequences from the LaRosa data set. We found that, in contrast to basic amino acids, acidic amino acids were rare in V3 sequences both within the consensus sequence and as sequence variants (data not shown). The predominant acidic amino acid substitutions were a conservative substitution (aspartic acid for glutamic acid) at position 25 and a change from glycine to either aspartic acid or glutamic acid at position 24.

Substitutions with uncharged amino acids were characterized by two patterns (Fig. 2B). First, at some positions there was a single amino acid that accounted for 40% or more of all substitutions at that position. Second, when these predominant amino acid polymorphisms were subtracted, V3 sequence heterogeneity was most significant at positions 13 and 20. The predominant amino acid at position 13 is histidine, with the substitutions representing small amino acids, while at position 20, the consensus phenylalanine is usually replaced with a hydrophobic amino acid. The next most heterogeneous position was position 25, which also had a frequent change to glutamine. These results, together with our analysis of charged substitutions, indicate that the apparent sequence variability in V3 has several distinct features, including substitution of basic amino acids at discrete positions, an underrepresentation of acidic amino acids, frequent substitutions that may represent polymorphisms, and clustered heterogeneity of uncharged amino acid substitutions at positions 13 and 20.

Sequence changes that correlate with altered tropism. Important determinants of macrophage tropism versus the ability to grow in T-cell lines have been shown to lie within the V3 loop (8, 9, 31, 63, 79), and the addition of basic amino acids to the V3 sequence enhances growth and syncytium formation in transformed T-cell lines (14, 15). It is likely that the predominant nonconservative basic amino acid substitutions at the positions seen in the examination of sequence heterogeneity (Fig. 2A) represent the naturally occurring positions for at least some of the sequence changes involved in tropism (positions 11, 24, 25, and 32). We tested the possibility that changes at these few positions could distinguish between macrophage-tropic and T-cell-line-tropic viruses by examining the V3 sequences of viruses with known phenotypes (Fig. 3A). All 10 of the T-cell-line-tropic viruses had a basic amino acid in at least one of these positions, and as with the larger data set, basic substitutions at these four positions accounted for approximately 75% of all nonconservative basic substitutions. In contrast, none of the 14 macrophage-tropic viruses had a basic substitution at any of the four positions. There was only one example of a nonconservative basic amino acid substitution at any position among the macrophage-tropic viruses (H13R in HIV-1 SF1703), while all V3 sequences from the group of T-cell-line-tropic viruses contained one or more nonconservative basic substitutions. The association of at least one nonconservative basic amino acid substitution with the T-cell-line-tropic viruses compared with the macrophage-tropic viruses was highly significant ($P < 10^{-4}$).

Comparison of V3 sequences from the two phenotypes also showed that all but three of the T-cell-line-tropic viruses had a nonacidic amino acid at position 25, while an acidic amino acid or alanine predominated at this position among macrophage-tropic viruses. Substitution of a nonacidic amino acid at position 25 among T-cell-line-tropic viruses, in contrast to the predominance of acidic amino acids at this position among macrophage-tropic viruses, was found to be significant ($P = 0.01$). Thus, three features that appear to distinguish T-cell-line-tropic from macrophage-tropic viruses include (i) the presence of nonconservative basic amino acid substitutions, (ii) frequent addition of basic amino acids at predominantly four positions, and (iii) a change from an acidic amino acid at position 25.

We next used these features to examine 10 pairs of V3 sequences derived from sequences present in vivo (Fig. 3B). These pairs of sequences represent two distinct groups of viral sequences present in each patient. In each case, there is one sequence with nonconservative basic substitutions which are frequently found at the positions noted among the viruses with the known T-cell-line-tropic phenotype. With one exception, the sequences in this group also had a nonacidic amino acid at position 25. The opposite was true for the second group of the paired sequences, in which position 25 was predominantly acidic and only one nonconservative basic amino acid change was present (Q32K in V12840). Thus, distinct features of viruses known to be T-cell-line tropic are also present within virus populations in vivo and appear to distinguish two different virus states in infected patients.

The link between the sequences in the larger data sets, which include sequences from cultured viruses, and the sequences present in vivo is strengthened by the observation that the consensus sequence of these in vivo-derived sequences is almost identical to the macrophage-tropic consensus sequence that was identified in the LaRosa data set, the only difference being an equal occurrence of glutamic acid and glutamine at position 25 (data not shown).

Discrete patterns of basic amino acid substitutions. Since the loss of an acidic amino acid at position 25 appeared to be an important determinant of T-cell-line tropism, we considered the possibility that the identity of the amino acid at this position affects the pattern of substitutions at other positions. We examined 46 V3 sequences with an uncharged glutamine at position 25 and compared these with 45 V3 sequences with a basic amino acid at position 25 (Fig. 4). The patterns of specific substitutions were considered as a function of increasing basic charge of the V3 loop. Overall, the patterns of sequence changes between these two groups appeared similar. However, there were two features that distinguished the two groups, both of which were statistically significant. First, with a basic amino acid at position 25, glycine replaced serine as the predominant uncharged amino acid at position 11 ($P < 10^{-4}$). Second, basic amino acid substitutions were found at position 13 only when a basic amino acid was present at position 25 ($P = 0.012$). Compar-

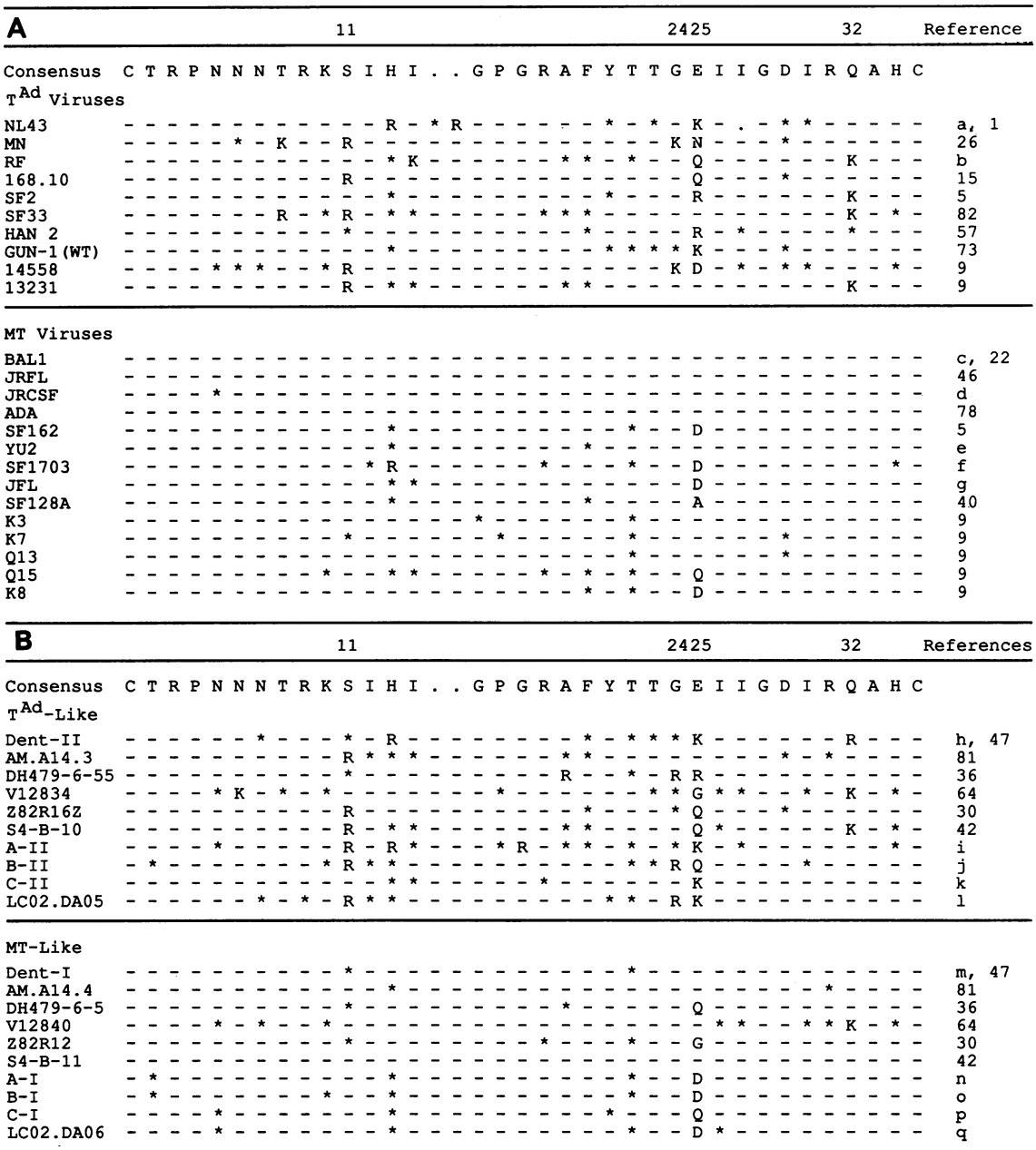

FIG. 3. Charge pattern of V3 sequences associated with known phenotypes and viruses present in vivo. Alignment with the V3 consensus sequence of (A) sequences from viruses known to grow in T-cell lines (top) or macrophages (bottom) and (B) sequences representing two distinct groups of viruses present in each of 10 HIV-1-infected patients at a single time point. These sequences represent a subset of the 83 V3 sequences from the HIV-1 data base (44) that are described in Materials and Methods. The sequences were obtained from uncultured PBMC DNA or plasma RNA. Each pair includes the predominant macrophage-tropic-like sequence (bottom) and a variant that contains the sequence elements characteristic of a T-cell-line-tropic virus (top). The pattern of basic charges and the amino acid present at position 25 were features used to sort the sequences in each pair into the two groups of viruses. The actual phenotype of these viruses is unknown. Amino acids shown represent nonconservative arginine and lysine substitutions. The identity of the amino acid at position 25 is also shown. Asterisks represent other substitutions from the consensus sequence. Identity with the consensus sequence is shown with a dash. Numbers at the top represent positions within the consensus sequence. T^{Ad}, T-cell adapted (i.e., T-cell-line tropic); MT, macrophage tropic. GenBank accession numbers: a, M19921; b, M12508; c, M68893; d, M38429; e, M93258; f, M66533; g, M31451; h, M90851; i, L21822; j, L21831; k, L21981; l, M90917; m, M90848; n, L21770; o, L21778; p, L21980; q, M90918.

ison of the basic amino acid substitution patterns for these two groups also showed that there was a tendency for arginine to be substituted at position 11 as the first basic amino acid substitution (outside of position 25) when glutamine rather than arginine or lysine was present at position 25. These results suggest that basic amino acid substitutions tend to follow at least two different pathways, although there are a few examples of sequences with elements of both

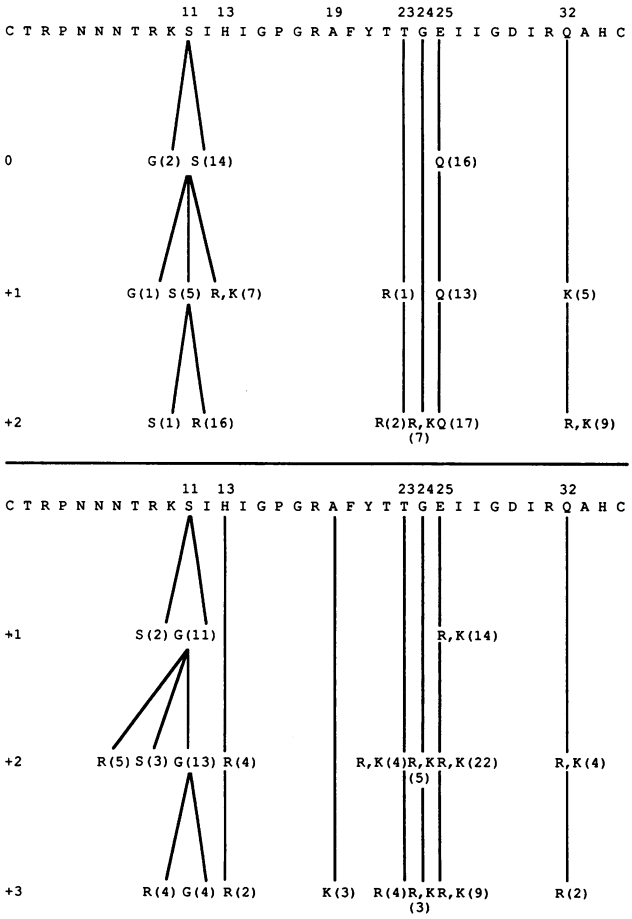

FIG. 4. Discrete patterns of basic amino acid substitutions within V3. Specific amino acid changes are shown for V3 sequences with either glutamine (top) or arginine or lysine (bottom) substituted at position 25. Substitution patterns are presented as a function of increasing basic charge in V3. Charges for individual V3 sequences were based on amino acid substitutions at seven positions within the LaRosa consensus sequence (11, 13, 19, 23, 24, 25, and 32). A total of 91 V3 sequences are represented, including 43 sequences from the LaRosa data set (see below) and 48 sequences from other sources (17, 21, 30, 44, 64): from reference 30, sequences C2, C3, D5, D7, and E3; from reference 21, sequences ACH-15.9, ACH-320, ACH-168.7, AMS-55, AMS-16.1, ACH-479.5, ACH-182.69, ACH-320.2A.5, and AMS-175; from reference 17, sequences W2-1a10, W2-1a6, W12-2c5, W1-1a1, and W1-1b5; from reference 44, isolates HIVWMJ12, HIVNY5NEW, HIVHAN, CANOB, and FO; from reference 64, sequences 91-a, 91-c, 82-c, and 87-d; and from the GenBank/EMBL data bases, accession numbers M90881, M90886, M74658, M26727, M90885, M21138, M21098, M38430, M90938, M90913, M90902, M90876, M90958, M90936, K02007, M90917, M90905, M74676, M90851, and M12508. The 43 sequences from the LaRosa data set include all sequences with a glutamine, arginine, or lysine residue at position 25 except for LaRosa V3 sequences 50, 106, 117, 126, 130, 146, 148, 149, 201, 203, 229, 235, 236, 238, and 240. A total of 132 nonconservative basic substitutions were found in the 91 V3 sequences, but in this analysis only the identity of the amino acids at positions 11, 13, 19, 23, 24, 25, and 32 were considered. Eighteen basic amino acid substitutions occurring at other positions within V3 are not shown. Amino acid substitutions are represented by a single letter. Shown in parentheses is the number of sequences with each substitution. Numbers at the top represent positions within the consensus sequence.

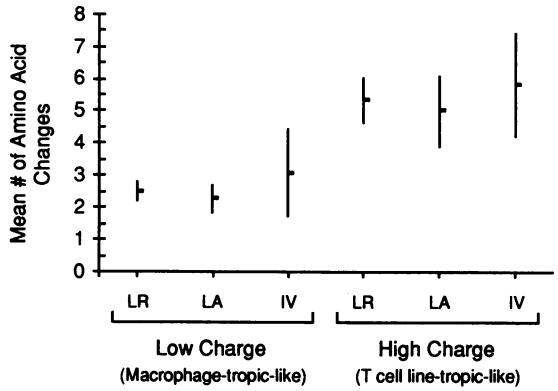

FIG. 5. Extent of sequence heterogeneity associated with distinct groups of V3 sequences. The mean number of amino acid differences from the V3 consensus sequence is shown for two groups of sequences in each of three separate data sets: the LaRosa data set (LR), 63 sequences from the HIV-1 data base (LA), and 10 pairs of in vivo-derived sequences (IV). The two groups represented include viruses that are macrophage-tropic-like and have a low net charge in V3 and sequence variants which have a high net charge in V3 and are T-cell-line-tropic-like. The V3 sequences in each data set were sorted into the two groups by the pattern of basic amino acid substitutions and the identity of the amino acid at position 25. The number of amino acid differences from the V3 consensus sequence, not including nonconservative basic substitutions, and the amino acid at position 25 was then determined for each sequence. Shown are the means for each group in the three data sets ±2 standard errors. V3 sequences from the LaRosa data set are described in the legend to Fig. 2. Sequences from the HIV-1 data base and the in vivo-derived sequences used in this figure are described in Materials and Methods and shown in Fig. 3B, respectively.

pathways. This indicates that while these patterns are preferred, they are not obligatory in the addition of basic amino acids to V3.

Basic substitutions and position 25 changes predict other sequence heterogeneity. From the previous analysis and the work of others (9, 14, 21, 79), we propose that the presence of basic amino acids predominantly at positions 11, 13, 19, 23, 24, and 32 of the V3 loop and the identity of the amino acids at position 25 can be used to distinguish two distinct sets of viral sequences. These features distinguish the sequences of viruses with known phenotypes as well as discrete subsets of viruses present in vivo (Fig. 3), although the effect on tropism of these individual substitutions in the macrophage-tropic background has not yet been tested. Nevertheless, we have used these features to sort V3 sequences from the three separate data sets (the LaRosa data set, the sequences from the HIV-1 data base, and the in vivo-derived sequences shown in Fig. 3B) into two groups.

Chesebro et al. have suggested that macrophage-tropic viruses are more similar in sequence as a group than T-cell-line-tropic viruses (9). We examined the extent of sequence heterogeneity associated with each group of viral sequences in each of the three data sets. As a measure of heterogeneity, we determined the number of amino acid differences from the consensus sequence for each sequence and did not include nonconservative basic substitutions in the tally.

The mean number of amino acid differences from the consensus sequence in each of the three data sets was approximately twofold greater among viruses that have basic amino acid substitutions, with approximately 2.6 changes as the mean for the macrophage-tropic-like viral sequences and

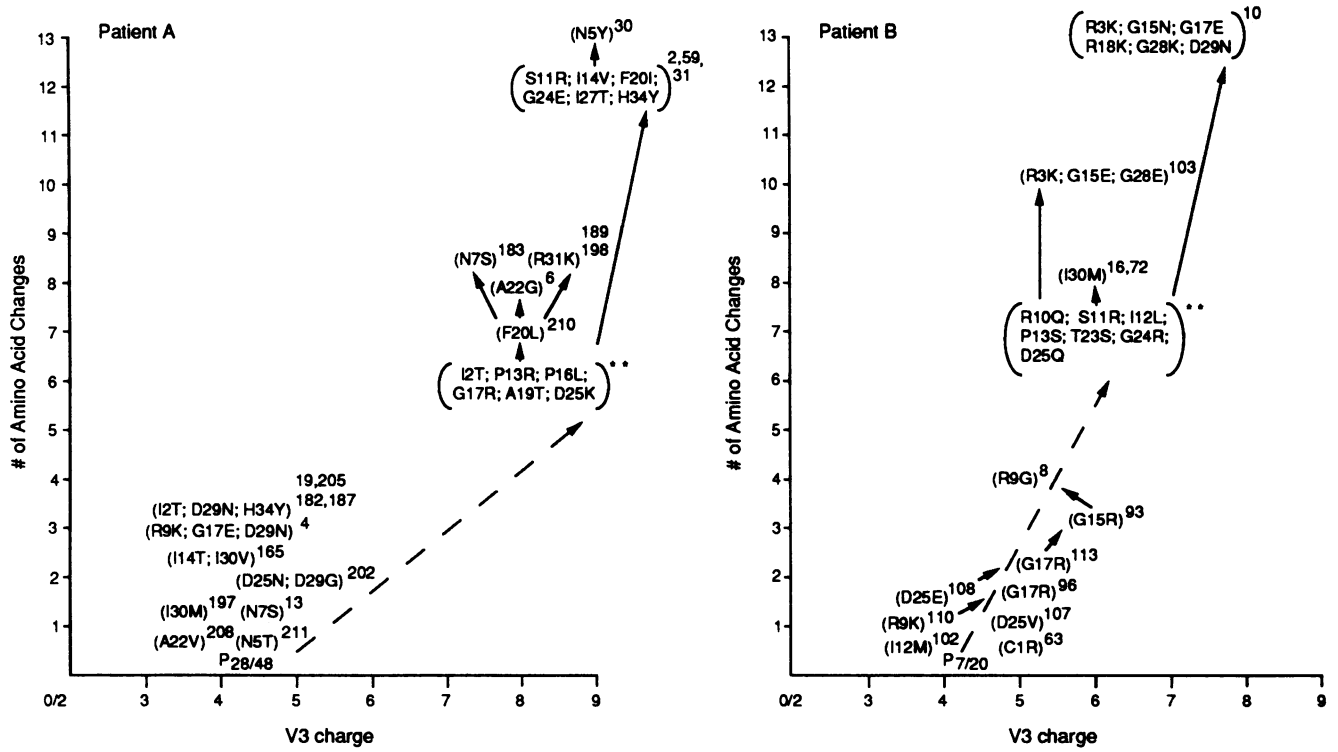

FIG. 6. Identification of discrete V3 sequence families in two HIV-1-infected patients. Individual V3 sequences from patients A and B are plotted as V3 charge versus total number of amino acid changes from the predominant V3 sequence in each patient. V3 charge for each clone was calculated from arginine and lysine substitutions and the amino acid residue at position 25. Represented are V3 sequences from 48 and 20 clones isolated at a single time point from patients A and B, respectively. The predominant or macrophage-tropic-like sequence in each patient, indicated by the letter P, is as follows: A, CIRPNNNTRKSIPIGPGRAFYATGDIIGDIRQAHC; B, CIRPNNNTRRSIPIGPGRAF YATGDIIGDIRQAHC. The subscript refers to the number of times that the predominant sequence appeared among the total V3 sequences examined in each patient. Superscript numbers refer to individual clone designations. Shown in parentheses are the amino acids that differ from the predominant sequence in each patient. The notation, e.g., A22V, includes the amino acid from the predominant V3 sequence, the position number, and the amino acid change present in a particular clone. Doubly starred superscripts represent an initial variant in each patient from which the highly divergent viruses must have evolved. Dashed arrows represent the monoclonal outgrowth of the initial variant from the predominant V3 sequence. V3 sequences at the end of solid arrows contain all of the changes found in the previous clone in addition to any new changes shown in parentheses. Nucleotide sequences for individual clones have GenBank accession numbers L21769 to L21836.

approximately 5.4 changes as the mean for the sequences that have basic substitutions (Fig. 5). The difference between the means for the two types of sequences represented in the in vivo data set was not as significant as the difference between the means for the larger data sets, possibly because of the smaller sample size. However, all three data sets followed the same trend. These results suggest that V3 sequences derived from putative macrophage-tropic viruses show about twofold less sequence heterogeneity than those sequences which contain basic amino acid substitutions.

Detection of discrete V3 sequence families *in vivo*. We used both the basic charge of the V3 loop and the total sequence variability to compare individual V3 sequences present in the virus population within an infected individual (Fig. 6). We examined a total of 68 V3 sequences amplified from PBMC DNA obtained from two HIV-1-infected patients (patients A and B) at a single time point. The most frequent (predominant) V3 sequence in each patient appeared to represent a sequence from a macrophage-tropic virus characterized by the presence of an acidic amino acid at position 25 and the absence of basic nonconservative amino acid substitutions (Fig. 6); in addition, the predominant sequence had a series of closely related sequences clustered around it. Both patient samples also showed evidence of highly divergent viral sequences. In these two patients, the pattern of V3 sequences among the divergent sequences shared three features. First, there was evidence for a monoclonal outgrowth of the distinct variant, as shown by the presence of several closely related sequences which were quite distinct from the predominant sequence. Second, each of the variants had a change at position 25 and at least one other basic substitution at position 11, 13, or 24. Third, in addition to the basic amino acid substitutions in V3, the variant V3 sequences had acquired numerous other sequence changes. These changes include substitutions of hydrophobic and small amino acids at the two most variable positions shown in Fig. 2B, position 20 (patient A; F20L) and position 13 (patient B; P13S), respectively. These results suggest that the acquisition of basic amino acids in vivo occurs when there is strong selective pressure for other sequence changes and that only discrete sequence variants accumulate.

DISCUSSION

V3 is a region within the HIV-1 *env* gene that displays significant sequence heterogeneity (37). In addition, this region influences the ability of the virus to replicate in macrophages and to grow in transformed T-cell lines (8, 9,

31, 63). V3 serves as a major target for neutralizing antibodies (25, 29, 32, 33, 34, 41, 45, 48, 53), suggesting that at least some of the heterogeneity may be the result of humoral selection. In this study, we have dissected V3 sequence variability within two large data sets of V3 sequences in order to separate sequence changes that correlate with virus tropism from more diverse changes that contribute to overall sequence heterogeneity. The consensus sequence, or average V3 sequence, is known to confer a macrophage-tropic phenotype. This sequence provided a baseline from which we were able to identify distinct features of the sequence variability associated with the V3 region.

Charge changes in V3 associated with the ability to replicate in transformed T cells. The addition of basic amino acids in V3 plays a role in the phenotypic shift between the NSI and the SI phenotypes (14, 15, 21). When we analyzed 175 published V3 sequences (37), we found very distinct patterns of basic amino acid substitutions. Nonconservative basic amino acid substitutions occurred predominantly at positions 11, 13, 19, 23, 24, 25, and 32 (Fig. 2A), with approximately 95% of all nonconservative basic substitutions either in or linked to a basic substitution at position 11, 24, 25, or 32. In addition, the presence of a basic or uncharged amino acid at position 25 correlated closely with basic amino acid substitutions at the other six positions. Thus, these two features of basic amino acid substitutions with a change at position 25 define a distinct subset of V3 sequences.

Comparison of 24 V3 sequences from viruses with known phenotypes provided strong evidence that the pattern of basic substitutions and the change at position 25 that we identified within the LaRosa data set are distinguishing features of a virus with a T-cell-line-tropic phenotype (Fig. 3A). With one exception, there were no nonconservative, basic amino acid substitutions present in the V3 sequences of 14 macrophage-tropic viruses compared with the consensus sequence. Among the group of 10 T-cell-line-tropic viruses, nonconservative basic substitutions were common and 75% of them occurred at the four predominant positions noted in our analysis of the 175 V3 sequences, i.e., 11, 24, 25, and 32.

It was also evident from comparing the identity of the amino acid at position 25 in the sequences of these two groups of viruses that the amino acid at this position was usually different for macrophage-tropic and T-cell-line-tropic viruses. Most of the macrophage-tropic viruses had either an acidic amino acid or alanine at position 25, in contrast to the T-cell-line-tropic viruses, which usually had a nonacidic amino acid at this position. Together, these observations strongly suggest that although the loss of an acidic amino acid at position 25 is a strong predictor of virus tropism, it is the combination of substitution of amino acids with a basic charge at a subset of residues in V3 (positions 11, 24, and 32) and the identity of the amino acid at position 25 that together can be used to distinguish between macrophage-tropic and T-cell-line-tropic viruses.

The importance of position 25 in predicting virus state has also been underscored previously in studies that used chimeric molecular clones and performed sequence comparisons of primary isolates. In a study by Westervelt et al. (79), amino acids in V3 associated with macrophage tropism included an acidic amino acid or alanine at position 25, while isolates incapable of infecting macrophages had a basic residue at this position and substitutions at positions 13 and 21. Fouchier et al. showed that among NSI viruses, the amino acid at position 25 was either acidic or uncharged and the amino acid residue at position 11 was uncharged (21). In contrast, either one or both of these residues were basic among SI isolates. Positions 11, 25, and 29 were shown by mutagenesis to contribute to an SI phenotype (14). Chesebro et al. (9) used molecular recombinants to show that sequence elements on either side of the GPGR motif can contribute to the T-cell-line-tropic phenotype. In examining a larger set of sequences, we have been able to generalize these patterns of sequence difference from the consensus sequence to allow a more consistent identification of two naturally occurring virus states.

Adaptive changes may follow several pathways. T-cell-line-tropic viruses can have V3 sequences with a single nonconservative basic amino acid substitution or up to three such substitutions (Fig. 3A). When we examined the positions of these substitutions as a function of the number of basic amino acid substitutions in the V3 sequence and whether the amino acid at position 25 was basic or an uncharged glutamine, we found several differences (Fig. 4). First, arginine appeared at position 13 only when there was a basic amino acid at position 25. Second, when there was a basic amino acid at position 25, glycine became the predominant uncharged amino acid at position 11 instead of serine. In this type of analysis, it is not possible to determine which change occurs first, although it is clear that the identities of the amino acids at positions 11 and 25 are coordinately influenced. Also, although these patterns represent strong tendencies, they are not unique, in that some sequences with elements of both patterns can be found. Nevertheless, the existence of these patterns must ultimately be understood in terms of the structural requirements for V3 function as reflected in either the direct or indirect interactions of these specific amino acids.

Some heterogeneity represents polymorphisms. When we examined uncharged amino acid substitutions within the consensus sequence, we found that a single amino acid represented a significant fraction of the substitutions away from the consensus sequence at many positions (Fig. 2B). For example, of 81 uncharged amino acid substitutions at position 22, 73 were from the consensus threonine to an alanine. Although we found no linkage between these predominant uncharged amino acid substitutions, the apparent selection for particular types of amino acid residues at specific positions further suggests that variability in V3 is limited by certain structural constraints. In a few cases, the polymorphism was most apparent in the viral sequences with nonconservative basic substitutions (A19V and E25Q), while in the rest of the cases, the polymorphisms appeared irrespective of the presence of nonconservative basic amino acid substitutions (Fig. 2B and data not shown).

Korber et al. found elements of sequence linkage after examining a large data set of V3 sequences (35). In the majority of cases, the most common amino acid combination noted represents the consensus sequence. However, there were examples among a small percentage of sequences in which the appearance of an amino acid at one site was linked to a particular amino acid at another position. These include R13 and K25 or R25; V19 and T13; V20 and K24 or R24; and G11 and H13 or R13.

Remaining sequence heterogeneity is partly clustered. When the basic amino acid substitutions, which are correlated with tropism, and the predominant, uncharged amino acid substitutions (i.e., polymorphisms) were subtracted from the total sequence heterogeneity in the LaRosa data set, we found that the remaining sequence heterogeneity was most significant at two positions, 13 and 20 (Fig. 2B). Both of these positions are two amino acids away from the GPGR

sequence. This sequence was initially modeled as a beta turn (37), a hypothesis that was subsequently confirmed by nuclear magnetic resonance analysis (4, 85). The initial modeling also suggested that a beta sheet conformation extends from the turn on each strand. In such a structure, the side chains at positions 13 and 20 would each be two residues away from the turn (on the N- and C-terminal sides, respectively), putting their side chains on the same face of the beta sheet and adjacent to each other across the sheet. Thus, these two variable residues may be equivalently placed to influence sequence heterogeneity, and therefore antigenicity, in the vicinity of the tip of the V3 loop.

Charge changes and heterogeneity are linked. When we sorted the published V3 sequences into two groups based on the parameters we have proposed to reflect the two virus states, we found that most of the sequence heterogeneity was associated with sequences that contain the nonconservative basic amino acid substitutions (Fig. 5). These findings were consistent for three separate data sets analyzed, including 10 pairs of V3 sequences present in vivo (Fig. 3B and 5). In all three data sets, the mean number of amino acid changes from the consensus sequence was approximately twofold greater among the group of sequences that contain basic amino acid substitutions than among the V3 sequences representing macrophage-tropic-like viruses.

Similar observations on the extent of sequence heterogeneity among viruses representing the two tropism phenotypes have been reported by others. Chesebro et al. noted more extensive sequence heterogeneity among T-cell-line-tropic viruses (9). Similarly, McNearney et al., in a study of V3 sequences obtained sequentially from infected patients, noted that V3 sequences obtained early in infection were more homogeneous than sequences obtained at later time points (42). This study also showed that the consensus sequences obtained from different patients at early time points were similar to each other and to the macrophage-tropic consensus sequence.

Although the significance of the additional sequence heterogeneity is not known, it is likely that the total sequence variability represents a sum of selective pressures. Early in infection, changes in the V3 sequence may be restricted, presumably because of some functional requirement, such as the establishment of HIV-1 infection in macrophages during transmission. As infection progresses, acquisition of basic charges may produce variants that have new properties (as evidenced by the ability to grow in transformed T-cell lines) in which restraints on the level of sequence change within V3 are relaxed. Alternatively, viruses with basic substitutions in V3 may replicate under conditions in which there is greater selection for sequence heterogeneity, a selection presumably supplied by the immune system.

Similar V3 sequence patterns are present *in vivo*. Basic substitutions seem to define a distinct subset of HIV-1 V3 sequence variants, although it is not clear whether all of the variants grow in transformed T-cell lines. Thus, sequence analysis can be used to complement the identification of this class of HIV-1 variants. We used this type of analysis to identify different sequence variants present in vivo (Fig. 3B). For each of 10 patients studied, two distinct groups of viruses were represented that could be readily distinguished from each other, in one case as a macrophage-tropic-like sequence and in the other by frequent substitution of basic amino acids at specific positions, and predominantly a nonacidic amino acid at position 25. In addition, these in vivo-derived sequences also displayed enhanced sequence heterogeneity among the sequences with basic amino acid substitutions (Fig. 5). Families of sequences representing these two states were also present among the virus populations in two patients whom we examined (Fig. 6). Thus, the major elements of both patterns of basic amino acid substitutions and sequence heterogeneity appear to be similar in each of the data sets examined.

Implication of V3 sequence heterogeneity. The function of V3 is not known, and at present, the V3 sequence can only be used as a genetic marker. In our analysis of V3 sequences, we have dissected sequence heterogeneity into three general classes: (i) patterns of basic amino acid substitutions whose presence is known to be correlated with improved virus replication in T cells (9, 14, 15, 21, 63); (ii) polymorphisms representing frequent specific substitutions not correlated with virus state; and (iii) residual heterogeneity that is linked to the presence of basic amino acid substitutions.

The observation of reduced sequence heterogeneity in the macrophage-tropic version of V3 isolated from diverse sources has two implications. First, it implies that sequences in this virus pool are separate from those in the T-cell-line-tropic viruses. One way to explain this observation is that the T-cell-line-tropic virus arises anew in each person from the macrophage-tropic virus. When we examined V3 sequences in vivo, we found that the T-cell-line-tropic sequences represented a discrete outgrowth of a variant with both charge changes and other sequence heterogeneity (Fig. 6). Thus, the increased heterogeneity associated with the T-cell-line-tropic viruses as a group may represent the sum of discrete viruses that evolve in different individuals rather than the appearance of many sequence variants in each individual.

The second implication that can be drawn is that it is the macrophage-tropic viruses that are most frequently transmitted and that usually establish the chronic infection. Recent examination of sequences from virus isolates obtained from five individuals prior to seroconversion showed that the V3 sequences were similar to each other and to those associated with a macrophage-tropic phenotype (83). McNearney et al. made a similar observation in a study of V3 sequences obtained early after infection from different patients (42). In each of these studies, the authors suggest that there is strong selection for a limited population of sequence variants during transmission and that only this subset of the HIV-1 quasispecies is capable of causing primary HIV infection. Transmission of a single virus species or the selective outgrowth of certain genotypes during primary infection was also suggested in a recent study of V3 sequences obtained at or soon after birth from 10 HIV-1-infected infants and their infected mothers (58). It is interesting that V3 sequences representing macrophage-tropic-like viruses appear to have been transmitted in all 10 of the mother-infant pairs studied even though T-cell-line-tropic-like variants were present in the virus pool of some of the mothers around the time of transmission. However, Roos et al. have shown that syncytium-inducing isolates can be transmitted (52), and in the case of transmission from the Florida dentist, both the dentist and one of the patients had two similar populations of V3 sequences which had recognizable sequence elements of macrophage-tropic and T-cell-line-tropic viruses (47). The persistence of both sequence variants in this patient suggests that the two virus types can be successfully transmitted, although the relative efficiency of transmission of each type of virus and the impact on disease course are not known.

ACKNOWLEDGMENTS

We thank Charles van der Horst, Robin Anderson, and Betsy Lopez for providing patient samples. We also thank Susan Fiscus and her staff for preparation of PBMCs. Additional thanks go to Bruce Chesebro and Bryan Cullen for helpful discussions.

This work was supported by Public Health Service grants U01-AI25868 and P01-CA19014.

REFERENCES

1. **Adachi, A., H. E. Gendelman, S. Koenig, T. Folks, R. Willey, A. Rabson, and M. A. Martin.** 1986. Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. J. Virol. **59**:284–291.
2. **Asjo, B., M. L. Morfeldt, J. Albert, G. Biberfeld, A. Karlsson, K. Lidman, and E. M. Fenyo.** 1986. Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection. Lancet **ii**:660–662.
3. **Cann, A. J., M. J. Churcher, M. Boyd, W. O'Brien, J. Q. Zhao, J. Zack, and I. S. Chen.** 1992. The region of the envelope gene of human immunodeficiency virus type 1 responsible for determination of cell tropism. J. Virol. **66**:305–309.
4. **Chandrasekhar, K., A. T. Profy, and H. J. Dyson.** 1991. Solution conformational preferences of immunogenic peptides derived from the principal neutralizing determinant of the HIV-1 envelope glycoprotein gp120. Biochemistry **30**:9187–9194.
5. **Cheng, M. C., M. Quiroga, J. W. Tung, D. Dina, and J. A. Levy.** 1990. Viral determinants of human immunodeficiency virus type 1 T-cell or macrophage tropism, cytopathogenicity, and CD4 antigen modulation. J. Virol. **64**:4390–4398.
6. **Cheng, M. C., D. Seto, M. Tateno, and J. A. Levy.** 1988. Biologic features of HIV-1 that correlate with virulence in the host. Science **240**:80–82.
7. **Cheng-Mayer, C., J. Homsy, L. A. Evans, and J. A. Levy.** 1988. Identification of human immunodeficiency virus subtypes with distinct patterns of sensitivity to serum neutralization. Proc. Natl. Acad. Sci. USA **85**:2815–2819.
8. **Chesebro, B., J. Nishio, S. Perryman, A. Cann, W. O'Brien, I. S. Chen, and K. Wehrly.** 1991. Identification of human immunodeficiency virus envelope gene sequences influencing viral entry into CD4-positive HeLa cells, T-leukemia cells, and macrophages. J. Virol. **65**:5782–5789.
9. **Chesebro, B., K. Wehrly, J. Nishio, and S. Perryman.** 1992. Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence heterogeneity in comparison with T-cell-tropic isolates: definition of critical amino acids involved in cell tropism. J. Virol. **66**:6547–6554.
10. **Clerici, M., D. R. Lucey, R. A. Zajac, R. N. Boswell, H. M. Gebel, H. Takahashi, J. A. Berzofsky, and G. M. Shearer.** 1991. Detection of cytotoxic T lymphocytes specific for synthetic peptides of gp160 in HIV-seropositive individuals. J. Immunol. **146**:2214–2219.
11. **Clerici, M., N. I. Stocks, R. A. Zajac, R. N. Boswell, D. C. Bernstein, D. L. Mann, G. M. Shearer, and J. A. Berzofsky.** 1989. Interleukin-2 production used to detect antigenic peptide recognition by T-helper lymphocytes from asymptomatic HIV-seropositive individuals. Nature (London) **339**:383–385.
12. **Coombs, R. W., A. C. Collier, J.-P. Allain, B. Nikora, M. Leuther, G. F. Gjerset, and L. Corey.** 1989. Plasma viremia in human immunodeficiency virus infection. N. Engl. J. Med. **321**:1626–1631.
13. **Crowl, R., K. Ganguly, M. Gordon, R. Conroy, M. Schaber, R. Kramer, G. Shaw, F. Wong-Staal, and E. P. Reddy.** 1985. HTLV-III env gene products synthesized in E. coli are recognized by antibodies present in the sera of AIDS patients. Cell **41**:979–986.
14. **De Jong, J.-J., A. De Ronde, W. Keulen, M. Tersmette, and J. Goudsmit.** 1992. Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. J. Virol. **66**:6777–6780.
15. **de Jong, J.-J., J. Goudsmit, W. Keulen, B. Klaver, W. Krone, M. Tersmette, and A. de Ronde.** 1992. Human immunodeficiency virus type 1 clones chimeric for the envelope V3 domain differ in syncytium formation and replication capacity. J. Virol. **66**:757–765.
16. **Dower, W. J., J. F. Miller, and C. W. Ragsdale.** 1988. High efficiency transformation of E. coli by high voltage electroporation. Nucleic Acids Res. **16**:6127–6145.
17. **Epstein, L. G., C. Kuiken, B. M. Blumberg, S. Hartman, L. R. Sharer, M. Clement, and J. Goudsmit.** 1991. HIV-1 V3 domain variation in brain and spleen of children with AIDS: tissue-specific evolution within host-determined quasispecies. Virology **180**:583–590.
18. **Evans, L. A., T. M. McHugh, D. P. Stites, and J. A. Levy.** 1987. Differential ability of human immunodeficiency virus isolates to productively infect human cells. J. Immunol. **138**:3415–3418.
19. **Fenyo, E. M., M. L. Morfeldt, F. Chiodi, B. Lind, A. von Gegerfelt, J. Albert, E. Olausson, and B. Asjo.** 1988. Distinct replicative and cytopathic characteristics of human immunodeficiency virus isolates. J. Virol. **62**:4414–4419.
20. **Fisher, R. A.** 1935. The logic of inductive inference. J. R. Stat. Soc. **98**:39–54.
21. **Fouchier, R. A., M. Groenink, N. A. Kootstra, M. Tersmette, H. G. Huisman, F. Miedema, and H. Schuitemaker.** 1992. Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. J. Virol. **66**:3183–3187.
22. **Gartner, S., P. Markovits, D. M. Markovitz, M. H. Kaplan, R. C. Gallo, and M. Popovic.** 1986. The role of mononuclear phagocytes in HTLV-III/LAV infection. Science **233**:215–219.
23. **Genetics Computer Group.** 1991. Program manual for the GCG package, version 7, April 1991. Genetics Computer Group, Madison, Wis.
24. **Goudsmit, J., N. K. Back, and P. L. Nara.** 1991. Genomic diversity and antigenic variation of HIV-1: links between pathogenesis, epidemiology and vaccine development. FASEB J. **5**:2427–2436.
25. **Goudsmit, J., C. Debouck, R. H. Meloen, L. Smit, M. Bakker, D. M. Asher, A. V. Wolff, C. J. J. Gibbs, and D. C. Gajdusek.** 1988. Human immunodeficiency virus type 1 neutralization epitope with conserved architecture elicits early type-specific antibodies in experimentally infected chimpanzees. Proc. Natl. Acad. Sci. USA **85**:4478–4482.
26. **Gurgo, C., H. G. Guo, G. Franchini, A. Aldovini, E. Collalti, K. Farrell, F. Wong-Staal, R. C. Gallo, and M. S. J. Reitz.** 1988. Envelope sequences of two new United States HIV-1 isolates. Virology **164**:531–536.
27. **Hahn, B. H., M. A. Gonda, G. M. Shaw, M. Popovic, J. A. Hoxie, R. C. Gallo, and F. Wong-Staal.** 1985. Genomic diversity of the acquired immune deficiency syndrome virus HTLV-III: different viruses exhibit greatest divergence in their envelope genes. Proc. Natl. Acad. Sci. USA **82**:4813–4817.
28. **Hahn, B. H., G. M. Shaw, M. E. Taylor, R. R. Redfield, P. D. Markham, S. Z. Salahuddin, F. Wong-Staal, R. C. Gallo, E. S. Parks, and W. P. Parks.** 1986. Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. Science **232**:1548–1553.
29. **Ho, D. D., M. G. Sarngadharan, M. S. Hirsch, R. T. Schooley, T. R. Rota, R. C. Kennedy, T. C. Chanh, and V. L. Sato.** 1987. Human immunodeficiency virus neutralizing antibodies recognize several conserved domains on the envelope glycoproteins. J. Virol. **61**:2024–2028.
30. **Holmes, E. C., L. Q. Zhang, P. Simmonds, C. A. Ludlam, and A. J. Brown.** 1992. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. Proc. Natl. Acad. Sci. USA **89**:4835–4839.
31. **Hwang, S. S., T. J. Boyle, H. K. Lyerly, and B. R. Cullen.** 1991. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. Science **253**:71–74.
32. **Javaherian, K., A. J. Langlois, G. J. LaRosa, A. T. Profy, D. P. Bolognesi, W. C. Herlihy, S. D. Putney, and T. J. Matthews.** 1990. Broadly neutralizing antibodies elicited by the hypervariable neutralizing determinant of HIV-1. Science **250**:1590–1593.

(Erratum, 251:13, 1991.)
33. **Javaherian, K., A. J. Langlois, C. McDanal, K. L. Ross, L. I. Eckler, C. L. Jellis, A. T. Profy, J. R. Rusche, D. P. Bolognesi, S. D. Putney, and T. J. Matthews.** 1989. Principal neutralizing domain of the human immunodeficiency virus type 1 envelope protein. Proc. Natl. Acad. Sci. USA **86**:6768–6772.
34. **Kenealy, W. R., T. J. Matthews, M. C. Ganfield, A. J. Langlois, D. M. Waselefsky, and S. R. J. Petteway.** 1989. Antibodies from human immunodeficiency virus-infected individuals bind to a short amino acid sequence that elicits neutralizing antibodies in animals. AIDS Res. Human Retroviruses **5**:173–182.
35. **Korber, B. T. M., R. M. Farber, D. H. Wolpert, and A. S. Lapedes.** Covariation of mutations in the V3 loop of HIV-1: an information theoretic analysis. Proc. Natl. Acad. Sci. USA, in press.
36. **Kuiken, C. L., J. J. de Jong, E. Baan, W. Keulen, M. Tersmette, and J. Goudsmit.** 1992. Evolution of the V3 envelope domain in proviral sequences and isolates of human immunodeficiency virus type 1 during transition of the viral biological phenotype. J. Virol. **66**:4622–4627. (Erratum, **66**:5704.)
37. **LaRosa, G. J., J. P. Davide, K. Weinhold, J. A. Waterbury, A. T. Profy, J. A. Lewis, A. J. Langlois, G. R. Dreesman, R. N. Boswell, P. Shadduck, L. H. Holley, M. Karplus, D. P. Bolognesi, T. J. Matthews, E. A. Emini, and S. D. Putney.** 1990. Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. Science **249**:932–935.
38. **LaRosa, G. J., J. P. Davide, K. Weinhold, J. A. Waterbury, A. T. Profy, J. A. Lewis, A. J. Langlois, G. R. Dreesman, R. N. Boswell, P. Shadduck, L. H. Holley, M. Karplus, D. P. Bolognesi, T. J. Matthews, E. A. Emini, and S. D. Putney.** 1991. Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant: corrections and clarifications. Science **251**:811.
39. **LaRosa, G. J., K. Weinhold, A. T. Profy, A. J. Langlois, G. R. Dreesman, R. N. Boswell, P. Shadduck, D. P. Bolognesi, T. J. Matthews, E. A. Emini, and S. D. Putney.** 1991. Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant: further clarifications. Science **253**:1146.
40. **Liu, Z. Q., C. Wood, J. A. Levy, and M. C. Cheng.** 1990. The viral envelope gene is involved in macrophage tropism of a human immunodeficiency virus type 1 strain isolated from brain tissue. J. Virol. **64**:6148–6153.
41. **Matsushita, S., G. M. Robert, J. Rusche, A. Koito, T. Hattori, H. Hoshino, K. Javaherian, K. Takatsuki, and S. Putney.** 1988. Characterization of a human immunodeficiency virus neutralizing monoclonal antibody and mapping of the neutralizing epitope. J. Virol. **62**:2107–2114.
42. **McNearney, T., Z. Hornickova, R. Markham, A. Birdwell, M. Arens, A. Saah, and L. Ratner.** 1992. Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. Proc. Natl. Acad. Sci. USA **89**:10247–10251.
43. **Modrow, S., B. H. Hahn, G. M. Shaw, R. C. Gallo, F. Wong-Staal, and H. Wolf.** 1987. Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: prediction of antigenic epitopes in conserved and variable regions. J. Virol. **61**:570–578.
44. **Myers, G., J. A. Berzofsky, B. Korber, R. F. Smith, and G. N. Pavlakis.** 1992. Human Retroviruses and AIDS. Theoretical Biology and Biophysics Group T-10, Los Alamos, N. Mex.
45. **Nara, P. L., R. R. Garrity, and J. Goudsmit.** 1991. Neutralization of HIV-1: a paradox of humoral proportions. FASEB J. **5**:2437–2455.
46. **O'Brien, W. A., Y. Koyanagi, A. Namazie, J. Q. Zhao, A. Diagne, K. Idler, J. A. Zack, and I. S. Chen.** 1990. HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120 outside the CD4-binding domain. Nature (London) **348**:69–73.
47. **Ou, C.-Y., C. A. Ciesielski, G. Myers, C. I. Bandea, C. C. Luo, B. T. Korber, J. I. Mullins, G. Schochetman, R. L. Berkelman, A. N. Economou, J. J. Witte, L. J. Furman, G. A. Satten, K. A. MacInnes, J. W. Curran, H. W. Jaffe, Laboratory Investigation Group, and Epidemiologic Investigation Group.** 1992. Molecular epidemiology of HIV transmission in a dental practice. Science **256**:1165–1171.
48. **Palker, T. J., M. E. Clark, A. J. Langlois, T. J. Matthews, K. J. Weinhold, R. R. Randall, D. P. Bolognesi, and B. F. Haynes.** 1988. Type-specific neutralization of the human immunodeficiency virus with antibodies to env-encoded synthetic peptides. Proc. Natl. Acad. Sci. USA **85**:1932–1936.
49. **Palker, T. J., T. J. Matthews, A. Langlois, M. E. Tanner, M. E. Martin, R. M. Scearce, J. E. Kim, J. A. Berzofsky, D. P. Bolognesi, and B. F. Haynes.** 1989. Polyvalent human immunodeficiency virus synthetic immunogen comprised of envelope gp120 T helper cell sites and B cell neutralization epitopes. J. Immunol. **142**:3612–3619.
50. **Rabson, A. B., and M. A. Martin.** 1985. Molecular organization of the AIDS retrovirus. Cell **40**:477–480.
51. **Ratner, L., A. Fisher, L. L. Jagodzinski, H. Mitsuya, R. S. Liou, R. C. Gallo, and F. Wong-Staal.** 1987. Complete nucleotide sequences of functional clones of the AIDS virus, HTLV-III. AIDS Res. Human Retroviruses **3**:57–69.
52. **Roos, M. T. L., J. M. A. Lange, R. E. Y. de Goede, R. A. Coutinho, P. T. A. Schellekens, F. Miedema, and M. Tersmette.** 1992. Viral phenotype and immune response in primary human immunodeficiency virus type 1 infection. J. Infect. Dis. **165**:427–432.
53. **Rusche, J. R., K. Javaherian, C. McDanal, J. Petro, D. Lynn, R. Grimaila, A. Langlois, R. C. Gallo, L. O. Arthur, P. J. Fischinger, D. P. Bolognesi, S. D. Putney, and T. J. Matthews.** 1988. Antibodies that inhibit fusion of human immunodeficiency virus-infected cells bind a 24-amino-acid sequence of the viral envelope, gp120. Proc. Natl. Acad. Sci. USA **85**:3198–3202.
54. **Saag, M. S., B. H. Hahn, J. Gibbons, Y. Li, E. S. Parks, W. P. Parks, and G. M. Shaw.** 1988. Extensive variation of human immunodeficiency virus type-1 in vivo. Nature (London) **334**:440–444.
55. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed., p. 1.92–1.99. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
56. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed., p. 10.60–10.61. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
57. **Sauermann, U., J. Schneider, J. Mous, U. Brunckhorst, I. Schedel, K. D. Jentsch, and G. Hunsmann.** 1990. Molecular cloning and characterization of a German HIV-1 isolate. AIDS Res. Human Retroviruses **6**:813–823.
58. **Scarlatti, G., T. Leitner, E. Halapi, J. Wahlberg, P. Marchisio, M. A. Clerici-Schoeller, H. Wigzell, E. M. Fenyo, J. Albert, M. Uhlen, and P. Rossi.** 1993. Comparison of variable region 3 sequences of human immunodeficiency virus type 1 from infected children with the RNA and DNA sequences of the virus populations of their mothers. Proc. Natl. Acad. Sci. USA **90**:1721–1725.
59. **Schuitemaker, H., M. Koot, N. A. Kootstra, M. W. Dercksen, R. E. de Goede, R. P. van Steenwijk, J. M. Lange, J. K. Schattenkerk, F. Miedema, and M. Tersmette.** 1992. Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytotropic to T-cell-tropic virus populations. J. Virol. **66**:1354–1360.
60. **Schuitemaker, H., N. A. Kootstra, R. E. de Goede, F. Miedema, and M. Tersmette.** 1991. Monocytotropic human immunodeficiency virus type 1 (HIV-1) variants detectable in all stages of HIV-1 infection lack T-cell line tropism and syncytium-inducing ability in primary T-cell culture. J. Virol. **65**:356–363.
61. **Schwartz, S., B. K. Felber, E. M. Fenyo, and G. N. Pavlakis.** 1989. Rapidly and slowly replicating human immunodeficiency virus type 1 isolates can be distinguished according to target-cell tropism in T-cell and monocyte cell lines. Proc. Natl. Acad. Sci. USA **86**:7200–7203.
62. **Shioda, T., J. A. Levy, and C. Cheng-Meyer.** 1991. Macrophage and T cell-line tropisms of HIV-1 are determined by specific regions of the envelope gp120 gene. Nature (London) **349**:167–169.
63. **Shioda, T., J. A. Levy, and C. Cheng-Mayer.** 1992. Small amino acid changes in the V3 hypervariable region of gp120 can affect

the T-cell-line and macrophage tropism of human immunodeficiency virus type 1. Proc. Natl. Acad. Sci. USA **89**:9434–9438.
64. **Simmonds, P., P. Balfe, C. A. Ludlam, J. O. Bishop, and A. J. Brown.** 1990. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. J. Virol. **64**:5840–5850.
65. **Simmonds, P., L. Q. Zhang, F. McOmish, P. Balfe, C. A. Ludlam, and A. J. Brown.** 1991. Discontinuous sequence change of human immunodeficiency virus (HIV) type 1 *env* sequences in plasma viral and lymphocyte-associated proviral populations in vivo: implications for models of HIV pathogenesis. J. Virol. **65**:6266–6276.
66. **Six, H. R., R. G. Webster, A. P. Kendal, W. P. Glezen, C. Griffis, and R. B. Couch.** 1983. Antigenic analysis of H1N1 viruses isolated in the Houston metropolitan area during four successive seasons. Infect. Immunol. **42**:453–458.
67. **Skinner, M. A., A. J. Langlois, C. B. McDanal, J. S. McDougal, D. P. Bolognesi, and T. J. Matthews.** 1988. Neutralizing antibodies to an immunodominant envelope sequence do not prevent gp120 binding to CD4. J. Virol. **62**:4195–4200.
68. **Starcich, B. R., B. H. Hahn, G. M. Shaw, P. D. McNeely, S. Modrow, H. Wolf, E. S. Parks, W. P. Parks, S. F. Josephs, R. C. Gallo, and F. Wong-Staal.** 1986. Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. Cell **45**:637–648.
69. **Takahashi, H., J. Cohen, A. Hosmalin, K. B. Cease, R. Houghten, J. L. Cornette, C. DeLisi, B. Moss, R. N. Germain, and J. A. Berzofsky.** 1988. An immunodominant epitope of the human immunodeficiency virus envelope glycoprotein gp160 recognized by class I major histocompatibility complex molecule-restricted murine cytotoxic T lymphocytes. Proc. Natl. Acad. Sci. USA **85**:3105–3109.
70. **Takahashi, H., R. N. Germain, B. Moss, and J. A. Berzofsky.** 1990. An immunodominant class I-restricted cytotoxic T lymphocyte determinant of human immunodeficiency virus type 1 induces CD4 class II-restricted help for itself. J. Exp. Med. **171**:571–576.
71. **Takahashi, H., S. Merli, S. D. Putney, R. Houghten, B. Moss, R. N. Germain, and J. A. Berzofsky.** 1989. A single amino acid interchange yields reciprocal CTL specificities for HIV-1 gp160. Science **246**:118–121.
72. **Takahashi, H., Y. Nakagawa, C. D. Pendleton, R. A. Houghten, K. Yokomuro, R. N. Germain, and J. A. Berzofsky.** 1992. Induction of broadly cross-reactive cytotoxic T cells recognizing an HIV-1 envelope determinant. Science **255**:333–336.
73. **Takeuchi, Y., M. Akutsu, K. Murayama, N. Shimizu, and H. Hoshino.** 1991. Host range mutant of human immunodeficiency virus type 1: modification of cell tropism by a single point mutation at the neutralization epitope in the *env* gene. J. Virol. **65**:1710–1718.
74. **Tersmette, M., R. A. Gruters, F. de Wolf, R. E. de Goede, J. M. Lange, P. T. Schellekens, J. Goudsmit, H. G. Huisman, and F. Miedema.** 1989. Evidence for a role of virulent human immunodeficiency virus (HIV) strains in pathogenesis of acquired immunodeficiency syndrome obtained from studies on a panel of sequential HIV isolates. J. Virol. **63**:2118–2125.
75. **Tersmette, M., R. E. de Goede, B. J. Al, I. N. Winkel, R. A. Gruters, H. T. Cuypers, H. G. Huisman, and F. Miedema.** 1988. Differential syncytium-inducing capacity of human immunodeficiency virus isolates: frequent detection of syncytium-inducing isolates in patients with acquired immunodeficiency syndrome (AIDS) and AIDS-related complex. J. Virol. **62**:2026–2032.
76. **Vieira, J., and J. Messing.** 1987. Production of single-stranded plasmid DNA. Methods Enzymol. **153**:3–11.
77. **Webster, R. G., A. P. Kendal, and W. Gerhard.** 1979. Analysis of antigenic drift in recently isolated influenza A (H1N1) viruses using monoclonal antibody preparations. Virology **96**:258–264.
78. **Westervelt, P., H. E. Gendelman, and L. Ratner.** 1991. Identification of a determinant within the human immunodeficiency virus 1 surface envelope glycoprotein critical for productive infection of primary monocytes. Proc. Natl. Acad. Sci. USA **88**:3097–3101.
79. **Westervelt, P., D. B. Trowbridge, L. G. Epstein, B. M. Blumberg, Y. Li, B. H. Hahn, G. M. Shaw, R. W. Price, and L. Ratner.** 1992. Macrophage tropism determinants of human immunodeficiency virus type 1 in vivo. J. Virol. **66**:2577–2582.
80. **Wiley, D. C., I. A. Wilson, and J. J. Skehel.** 1981. Structural identification of the antibody-binding sites of Hong Kong influenza hemagglutinin and their involvement in antigenic variation. Nature (London) **289**:373–378.
81. **Wolfs, T. F., G. Zwart, M. Bakker, and J. Goudsmit.** 1992. HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. Virology **189**:103–110.
82. **York-Higgins, D., C. Cheng-Mayer, D. Bauer, J. A. Levy, and D. Dina.** 1990. Human immunodeficiency virus type 1 cellular host range, replication, and cytopathicity are linked to the envelope region of the viral genome. J. Virol. **64**:4016–4020.
83. **Zhang, L. Q., P. MacKenzie, A. Cleland, E. C. Holmes, A. J. L. Brown, and P. Simmonds.** 1993. Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. J. Virol. **67**:3345–3356.
84. **Zoller, M. J., and M. Smith.** 1983. Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors. Methods Enzymol. **100**:468–500.
85. **Zvi, A., R. Hiller, and J. Anglister.** 1992. Solution conformation of a peptide corresponding to the principal neutralizing determinant of HIV-1_{IIIB}: a two-dimensional NMR study. Biochemistry **31**:6972–6979.