LA-UR--91-4102

DE92 004579

TITLE HIV-1 SEQUENCE VARIATION BETWEEN ISOLATES FROM MOTHER-INFANT TRANSMISSION PAIRS

AUTHOR(S): CARLA M. WIKE
MICHAEL R. DANIELS
MANOHAR FURTADO
M. WOLINSKY
BETTE KORBER
CECELIA HUTTO
JOSE MUNOZ
WADE PARKS
ALFRED SAAH

SUBMITTED TO For publication in a special edition of AIDS RESEARCH AND HUMAN RETROVIRUSES. PROCEEDINGS FROM THE INTERNATIONAL CONFERENCE ON ADVANCES IN AIDS VACCINE DEVELOPMENT: FOURTH ANNUAL MEETING OF THE NATIONAL COOPERATIVE VACCINE DEVELOPMENT GROUP FOR AIDS

DISCLAIMER

Los Alamos Los Alamos National Laboratory
Los Alamos, New Mexico 87545

HIV-1 SEQUENCE VARIATION BETWEEN ISOLATES FROM MOTHER-INFANT TRANSMISSION PAIRS

Carla M. Wike[1], Bette Korber[2], Michael R. Daniels[1], Cecilia Hutto[3], Jose Munoz[4], Manohar Furtado[1], Wade Parks[4], Alfred Saah[5], and Steven M. Wolinsky[1].

[1]Department of Medicine, Northwestern University Medical School, Chicago, IL 60611; [2]Theory Division, T-10, Los Alamos National Laboratories, Los Alamos, NM 87545; [3]Department of Pediatrics, University of Miami School of Medicine, Miami, FL 33136; [4]Department of Pediatrics, New York University School of Medicine, New York, NY 10016; [5]Department of Epidemiology, Johns Hopkins School of Hygiene and Public Health, Baltimore, MD 21205.

ABSTRACT

To examine the sequence diversity of human immunodeficiency virus type 1 (HIV-1) between known transmission sets, sequences from the V3 and V4-V5 region of the env gene from 4 mother-infant pairs were analyzed. The mean interpatient sequence variation between isolates from linked mother-infant pairs was comparable to the sequence diversity found between isolates from other close contacts. The mean intrapatient variation was significantly less in the infants' isolates then the isolates from both their mothers and other characterized intrapatient sequence sets. In addition, a distinct and characteristic difference in the glycosylation pattern preceding the V3 loop was found between each linked transmission pair. These findings indicate that selection of specific genotypic variants, which may play a role in some direct transmission sets, and the duration of infection are important factors in the degree of diversity seen between the sequence sets.

INTRODUCTION

The complex distribution of human immunodeficiency virus type 1 (HIV-1) genotypic variants, or quasispecies, within infected individuals has been characterized extensively (1). The wide range of diversity reflects variation in both geographic distribution and time of sampling for both linked and unlinked patients, respectively.

To characterize sequence diversity within a well defined context of transmission, we studied the quasispecies populations found in mother-infant transmission pairs for the V3 and V4-V5 regions. These data were compared with similar sequence data from isolates from both the Edinburgh hemophilia cohort and an infected Long Island family (2-5). Our results indicate that the range of diversity between isolates from all these epidemiologically linked sequence sets may be a function of the mode of transmission, the time of sampling following infection, and selection for change within the host.

MATERIALS AND METHODS

Data for this study was obtained from four maternal-infant transmission pairs and the sequence sets summarized in the Human Retroviruses and AIDS database, 1991. To characterize the distinct genotypic variants from the mother-infant transmission pairs, DNA

extracted directly from peripheral blood mononuclear cells (PBMCs) of four mothers and their infants was amplified by PCR using nested primer pair sets. The primer sets flanked both the V3 region containing the immunodominant loop, and the V4-V5 regions which encompasses a defined part of the conserved CD4-binding domain. Three of the mother-infant pairs were from the United States; two of the three mothers are of Haitian descent currently residing in Florida. The fourth mother-infant pair was from Rwanda; only V3 sequences were available from this pair. Product DNA from 5 to 27 individual samples derived from transformed bacterial colonies was sequenced from each patient in both directions. Sequences were analyzed using the MASE program for alignment and obtaining similarity scores (6). Similarity analyses were done discounting positions in which gaps were inserted to maintain the alignment. The proportion of synonymous to nonsynonymous substitutions was calculated by the method of M. Nei and T. Gojobori (7).

RESULTS

Sequence Variation. Intrapatient analysis showed extensive diversity within the mother's HIV-1 sequences. The mothers' isolates were characterized by numerous amino acid deletions, insertions, and substitutions, especially in the V4-V5 regions. The infants' isolates were more highly conserved. Similarity analysis showed that the mothers' viral sequences varied by up to 15% in the V4-V5 region, while the infants' sequences varied at most by only

4%. For the V3 region, the mothers' sequences showed a similar range of variability as the V4-V5, with sequences that varied by up to 15%, but the distribution was different. In the infant V3 sequences, a few sequences varied by up to 15%, but the majority of the sequences varied by less than 4%. While the median distance between intrapatient V3 sequences for all four of the mothers' sequence sets was 5%, the infants' sequence sets were more conserved with a median distance of 2.5% between sequences.

Interpatient sequence analysis of the three mother-infant pairs from Haiti and the U.S. (mother-infant pair 4, from Rwanda was not included in this analysis) showed a wide range of diversity between unlinked mothers and infants, while interpatient variation between related mothers and their infants was markedly less. The median interpatient sequence variation between mothers and their infants was 6.2% (range: 0% to 28%) for the V3 region, and 4.4% (range: 0.3% to 15%) in the V4- V5 regions. The median V3 variation between unlinked mothers was 15.4% (range: 9% to 25%) and the median V4-V5 variation was 12% (range: 7.9% to 17.2%).

Synonymous to Nonsynonymous Substitutions. Within the mother-infant data set, nonsynonymous substitutions predominated over synonymous substitutions in the mother's V3 and V4-V5 domains, with ratios varying among individuals. Within the V3 region from the infants isolates, the ratios approximated 1, while the V4-V5 region the infant's ratios approximated 2. The lower ratio for the

infant's V3 region compared to the V4-V5 region is consistent with the possibility that the V3 region is under greater pressure for change and is evolving at a more rapid rate. This disproportionate rate of evolution is also indicated by the presence of more divergent sequences in the V3 than the V4-V5 regions in the infants' sequence sets. In the V3 region, the average proportion of synonymous to nonsynonymous substitutions in pairwise comparisons of all intrapatient sequences, $\langle P_s/P_n \rangle$, was 0.61, 1.90, 0.90, and 1.2 for the 4 mothers' sequences and 1.3, 0.2, 1.1, and 0.88 for their infant's sequences. In the V4-V5 domains, synonymous substitutions were generally less common than nonsynonymous substitutions in the three mothers' sequence sets with $\langle P_s/P_n \rangle$ ratios of 0.70, 1.10, and 0.83. In the corresponding regions of the infants' sequences, synonymous substitutions predominated with $\langle P_s/P_n \rangle$ ratios of 2.0, 1.7, and 2.2 (only V3 sequences are available for mother-infant pair 4).

N-linked glycosylation sites. A survey of the N-linked glycosylation sites in the region of the V3 loop from the mother-infant transmission sets and selected North American, European, and African sequence isolates, showed several interesting features. An N-linked glycosylation site proximal to the first conserved cysteine of the V3 loop present in each of the four mother's sequence sets, was completely absent from all sequences in the four infants' sequence sets. In the case of mother1 (mom1), this site was present in only a small number of sequences (2/9), hence it

does not appear in the consensus shown in Fig. 1. This glycosylation site was present in the majority of the mothers' sequences, comprising 18/19, 19/22, and 17/23 of the sequences for mothers 2, 3, and 4, respectively. This site is also highly conserved within the other isolates in the database (present in 27/29) except in the African sequences ELI and Z321, which similar to the infants, lack this site (Fig. 1). The N-X-S sequon proximal to the first cysteine of the V3 loop (at amino acid position 289 to 291 relative to LAI) in mother-infant sets 1 and 4, as well as African isolates U455 and Z321 are all missing the asparagine residue (N) of this sequon. Four out of the 7 individuals lacking this glycosylation site have an unusual glycosylation site appearing just upstream of the original site (mom1, infant1, infant4 and Z321).

DISCUSSION

Limited sequence data exist which characterize the HIV-1 genotypic variation found between close contacts soon after transmission. To examine the sequence relationships between epidemiologically linked and unlinked patients, the mother-infant data was compared to the Edinburgh Cohort and the Long Island family set. The various data sets were evaluated for the extent of sequence variation, rates of synonymous and nonsynonymous substitutions, and the N-linked glycosylation patterns which precede the V3 loop to examine the diversity between the related and non-related sequence sets.

The median interpatient nucleotide sequence distance for the linked mother-infant transmission pairs was 6.2% for the V3 (range: 0% to 28%) and 4.4% for the V4-V5 domains (range: 0.3% to 14.6%). The V3 sequences of the Rwandan mother-infant pair are more divergent. For the Long Island family sequence set, the mean interpatient nucleotide sequence distance for the V3 and partial V4-V5 region was 8.5% between the mother and her child and 3.7% between the mother and her partner. The mean sequence distance between members of the Edinburgh hemophiliac cohort ranged between 9.4% and 14.8% for V3 and 5.6% and 11.1% for V4-V5.

Although the extent of the regions being compared is generally equitable, the trend towards increased variability with time is noteworthy. The mean sequence differences found between the linked

mother-infant pairs, the members of the Edinburgh hemophiliac cohort, and mother-partner from the Long Island family sequence set confirms the close relationship that exists between temporally related, epidemiologically linked sets. The greater diversity between the mother from Long Island and her child, both of whom were infected for over 11 years before sampling, than between the mother and her partner, for whom the duration of infection is not known, suggest that significant evolution of the genotypic variants occurred. The sequence sets from the Edinburgh Cohort also show that epidemiologically linked patients can diverge significantly over time.

The intrapatient sequence variation in the V4-V5 domains for the mothers isolates ranged up to 15% while their infants isolates varied by no more than 4%. Similarly, the intrapatient sequences from the V3 region were more conserved in the infants than their mothers with the majority of the infants' isolates varying less than 4%. The mothers' V3 and V4-V5 median sequence difference compares with the mean intrapatient nucleotide sequence distances of 4.2% for the V3 and 5.5% for the V4-V5 regions as calculated by Balfe et al. for the Edinburgh cohort (5). Since a direct sequencing technique was used to characterize the sequence differences for the Long Island family, the range of intrapatient sequence variation can not be assessed.

The intrapatient differences between the two sequence sets can be

attributed to both the mode and timing of transmission. Following perinatal transmission, a bottleneck phenomena may occur whereby a single genotypic variant may escape the host's surveillance mechanisms. This would account for the initial narrow distribution of variants, followed later by an expansion and evolution of the virus in the host in response to differing selection constraints. Similar evolution would explain the differences found for the hemophiliacs, although infection by transfusion could involve a greater and more diverse inoculum than vertical transmission.

The high rate of nucleotide sequence variation may or may not conserve the integrity of the sequence at the amino acid level. Synonymous substitutions which are usually well tolerated in regions critical for structure and function, result in a conservation of an amino acid, while nonsynonymous substitutions result in a non-conserved change. A high rate of nonsynonymous substitutions has been shown to be a correlate of positive Darwinian selection in several systems with known biological relevance (8). In the absence of selection, the ratio of synonymous to nonsynonymous substitutions would be one. Because of structural and functional constraints on the evolution of proteins, this ratio is typically greater than one. If positive selection for change at the amino acid level dominates, the ratio can fall below one. An example of this is within the HLA class I and II molecules: the specific nucleotides encoding the residues in the antigen recognition sites have a low ratio (significantly less than

1), the exon they are located in has ratio around 1, and the other exons of the molecule have ratios of 5 to 6. The selective advantage of the variability of the antigen recognition sites is presumably to allow interaction with a large repertoire of epitopes. In the case of the V3 loop we are averaging over a whole region, in which some residues are highly variable and others are very conserved (1). Patterns of specific variable residues can be observed in individuals, which may be the result of host specific selection pressure on specific residues due to the B or T cell repertoire of that individual.

The ratio of synonymous to nonsynonymous substitutions were calculated to determine the strength of selection on existing amino acid sequences. For the mother-infant transmission pairs, nonsynonymous substitutions tended to predominate over synonymous substitutions. The total number of amino acid changes were greater in the mothers' viral sequences than in their infants' sequences. Within the Edinburgh Cohort the frequency of nonsynonymous to synonymous substitutions for the V4-V5 and V3 regions were comparable to the ratios for the mothers. The low ratios, even when synonymous substitutions predominated, indicate that the gp120 hypervariable domains are under enormous selection pressure for change. This type of analysis is limited, however, when comparing small sets of relative short sequences, since the statistical significance is low when considering a small number of changes. As the reference sets and number of observed changes increase, the

statistical significance increases and allows a more accurate estimation of evolutionary relationships.

Co-translational and postranslational carbohydrate additions to N-linked glycosylation sites (N-X-T or N-X-S sequons) can contribute to the formation of conformational epitopes and obscure peptide epitopes. Both of these changes may facilitate the viruses ability to evade the host's humoral and cell-mediated immune responses (9-13). The striking absence of an N-linked glycosylation site (amino acid site 295) in all four of the infants may provide evidence for the evolution and selection of a particularly strain for transmission. Alterations in the glycosylation pattern preceding the V3 loop (at position 295 relative to LAI) for the partner in the Long Island family also suggest that changes in this region are important for some routes of transmission.

In summary, these results further demonstrate the extreme complexity and heterogeneity of HIV-1 which exists within and between infected patients. Furthermore, the degree of diversity may relate to the route of transmission, the length of time for which the patient was infected, and selection pressures within the host.

ACKNOWLEDGMENTS

This work was supported by grants from the NIH, NICHD and NIAID.

REFERENCES

1. Myers G, Korber B, Berzofsky J, Smith R, and Pavlakis GN, eds. 1991. Human retroviruses and AIDS, a compilation and analysis of nucleic acid and amino acid sequences, published by Los Alamos National Laboratory, T-10, Los Alamos, NM.

2. Wolinsky SM, Wike CM, Korber B, Hutto C, Parks W, Rosenblum L, Kunstman K, Furtado M, and Munoz J: Selective transmission of human immunodeficiency virus type 1 (HIV-1) variants from mother to infants. Science, in press.

3. Burger H, Weiser B, Flaherty K, Gulla J, Nguyen P, and Gibbs RA: Evolution of human immunodeficiency virus type 1 nucleotide sequence diversity among close contacts. Proc Natl Acad Sci (USA), in press.

4. Simmonds P, Balfe P, Ludlam CA, Bishop JO, and Leigh Brown AJ: Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. J Virol 1990;64:5840-5850.

5. Balfe P, Simmonds P, Ludlam CA, Bishop JO, and Leigh Brown AJ: Concurrent evolution of human immunodeficiency virus type 1 in patients infected from the same source: rate of sequence change and low frequency of inactivating mutations. J Virol 1990;64:6221-6233.

6. Sequences were analyzed using the MASE program (D. Falkner and Y. Jerka, MBCRR, Harvard University) for alignment similarity scores. Analyses were done discounting positions in which gaps were inserted to maintain the alignment.

7. Li WH, Wu CI, and Luo CC: A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon change. Mol Biol Evol 1985;2(2):150-174.

8. Hughes AL, and Nei M: Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc Natl Acad Sci (USA) 1989;86:958-962.

9. Alexander S and Elder JH: Carbohydrate dramatically influences immune reactivity of antisera to viral glycoprotein antigens. Science 1984;226:1328-1332.

10. Sodora DL, Cohen GH, and Eisenberg RJ: Influence of asparagine-linked oligosaccharides on antigenicity, processing, and cell surface expression of herpes simplex virus type 1 glycoprotein d. J Virol 1989;63:5184-5193.

11. Skehel JJ, Stevens DJ, Daniels RS, Douglas AR, Knossow M, Wilson IA, and Wiley DC: A carbohydrate side chain on hemogglutinins of Hong Kong influenza viruses inhibits recognition

by a monoclonal antibody. Proc. Natl Acad Sci (USA) 1984;81:1779-1783.

12. Feizi T and Childs RA: Carbohydrates as antigenic determinants of glycoproteins. Biochem J 1987;245:1-12.

13. Jones IM and Jacob GS: Anti-HIV drug mechanism. Nature 1991;352:198.

Address reprint requests to: Dr. Steven M. Wolinsky, Department of Medicine, Northwestern University Medical School, Olson 8427, 710 N. Fairbanks, Chicago, IL 60611.

FIGURE LEGEND

Fig.1. Amino acid sequence alignments from the V3 loop region from mother-infant pairs and North American, European, and African sequences currently in the AIDS and Human Retroviruses Database. The 50% consensus sequence for each mother and infant (mom-tot) and the consensus sequences of characteristic North American, European, and African sequences (ELI, Z2Z6, NDK, JY1, MAL, U455, Z321) has been aligned with the North American/European/African consensus sequence. Amino acids which match the consensus at the top of the alignments are indicated by a "-". Deletions are indicated by a ".". Above the consensus are a group of symbols. The "***" indicate the most highly conserved N-linked glycosylation sites, and the "**" indicates the highly conserved cysteine involved in the formation of the V3 loop. This is a fragment of the region sequenced.

```
                                  *
                              ***   ***   ***
CONSENSUS           IVQLNESVEINCTRPNNNTRKSI
mom1.CONSENSUS      ----QTP-N-T------------
tot1.CONSENSUS      ----QTP-N-T------------
mom2.CONSENSUS      ---?-T-----------------
tot2.CONSENSUS      -----T--D-T------------
mom3.CONSENSUS      ------T-V--------------
tot3.CONSENSUS      ------T---D-I----------
mom4.CONSENSUS      ----TKP-K---------I---V
tot4.CONSENSUS      ----AKP-N-T-I----------
HIVLAI              -----Q-----------------
HIVHXB2R            -----T---------------R-
HIVNL43             -----T-----------------
HIVMFA              -----T-----------------
HIVMN               --H-----Q-------Y-K--R-
HIVBRVA             ---------------------R-
HIVSC               ----K-A------------TR--
HIVJH3              ----K-P-V------SKT--RR-
HIVALA1             ----------------IYRKGR-
HIVBAL1             -----------------------
HIVJRCSF            --------K------S-------
HIVJRFL             ----K------------------
HIVOYI              -----K--------------NR-
HIVSF2              --------A--------------
HIVNY5CG            ---------------------G-
HIVSF162            ----K------------------
HIVCDC4             -----V-----------H---RV
HIVSF33             L----V------------R-RR-
HIVHAN              --H------------------G-
HIVADA              ----K------------------
HIVWMJ2             --H------------Y--V-R-L
HIVRF               -----A--Q--------------
HIVELI              -AH-----K-T-A--YQ---QRT
HIVZ2Z6             --------A------YR-I-QRT
HIVNDK              -----A-IV------YKY--QRT
HIVJY1              --H------------D-KITRQS
HIVMAL              ------T-T------G----RG-
HIVU455             ----VNP-K---S--Y.----N-
HIVZ321             ----VKP-N-T-M----------
```