Sequence Analysis of the Dimerization Initiation Site of Concordant and Discordant Viral Variants Superinfecting HIV Type 1 Patients

Luzia Mayr,¹ Rebecca Powell,² Thompson Kinge,³ and Phillipe N. Nyambi^{1,4}

Abstract

For HIV recombination to occur, the RNAs from two infecting strains within a cell must dimerize at the dimerization initiation site (DIS). We examined the sequence identity at the DIS (697–731 bp, Hxb2 numbering engine) in patients superinfected with concordant HIV-1 strains and compared them to those with discordant strains. Viral RNA in sequential plasma from four subjects superinfected with subtype-discordant and two subjects superinfected with subtype-concordant HIV-1 strains was extracted, amplified (5' LTR-early gag: 526–1200 bp, Hxb2 numbering engine), sequenced, and analyzed to determine their compatibility for dimerization *in vivo*. The concordant viruses infecting the two subjects exhibited identical sequences in the 35-bp-long DIS region while sequences from the discordant viruses revealed single nucleotide changes that were located in the DIS loop (715 bp), its flanking nucleotides (710 bp and 717 bp), and the DIS stem (719 bp). Evidence from *in vitro* experiments demonstrates that these *in vivo* changes identified can abolish dimerization and reduce recombination frequency. Therefore, these results revealing differences in the DIS of discordant strains versus the similarity noted for the concordant strains may contribute to the differences in the frequency of recombination in patients superinfected with such HIV-1 variants.

THE HIV-1 PANDEMIC IS COMPOSED of a wide array of genetically diverse strains including innumerable recombinant forms.¹ This broad genetic diversity is in part generated by superinfection of one cell by two or more genetically discordant viruses, whose genomes recombine during reverse transcription, producing new mosaic virus strains that can then spread in the population.^{2,3} For recombination to happen, the RNA genomes of superinfecting viruses have to dimerize at the dimerization initiation site (DIS), a 35-nucleotide sequence that is located in the 5' untranslated region of HIV-1 (697–731 bp, Hxb2 numbering engine). The DIS of each viral RNA folds into a hairpin structure that then interact, forming a loop–loop kissing dimer. HIV-1 recombination results from reverse transcriptase template switching between the two dimerized viral RNAs.^{4,5} We recently conducted a longitudinal study in which we analyzed three genetic loci (the C1C2 region of *env*, RT region of *pol*, and *vif* accessory gene region) of HIV-1 in superinfected subjects.⁶ No recombinants or recombination breakpoints were detected in the sequences analyzed of the subtype-discordant infecting strains. On the

contrary, several recombinants with characteristic breakpoints were detected in subjects superinfected with subtype-concordant strains.⁶

An important factor that affects HIV recombination is the extent of sequence similarity among two viral strains, especially compatibility of the sequences at the DIS.^{7,8} Indeed, it was demonstrated *in vitro* that intersubtype recombination events occur at a much lower rate than intrasubtype recombinations. The use of HIV-1 subtype B and C-based vectors showed that nucleotide differences in the DIS region account for a reduction of intersubtype recombination while matching of the DIS sequences resulted in a 4-fold elevation of the B/C intersubtype recombination rate.⁸ Similar results were obtained with HIV-1 5' RNA sequences whereby deletion of the entire DIS as well as single point mutations abolished *in vitro* dimerization, while the introduction of compensatory mutations restored the process.⁹ Certain nucleotides such as the central GC pair of the loop palindrome as well as adenine residues that normally flank the palindrome seem to be critical for dimerization initiation.¹⁰ Thus far, no study has

¹Department of Pathology, ²Department of Microbiology, New York University School of Medicine, New York, New York.

³Ministry of Public Health, Yaoundé, Cameroon.

⁴Manhattan Veterans Affairs Harbor Healthcare Systems, New York, New York.

examined the DIS in patients superinfected with HIV-1 strains and how this could possibly affect the frequency of generation of recombinants.

The aim of our study was to sequence and analyze the DIS in patients superinfected with subtype-discordant ($n=4$) and subtype-concordant ($n=2$) HIV-1 strains to determine their identity and compatibility for dimerization *in vivo*. Samples were collected from antiretroviral drug-naïve patients in Yaoundé, Cameroon in calendar years 2001–2004. The patients studied included CMNYU107, CMNYU124, CMNYU129, CMNYU6544, CMNYU6506, and CMNYU6564. Sequential blood samples containing two different viral strains were taken at intervals of 3 months (CMNYU6564), 9 months (CMNYU124, CMNYU129), 12 months (CMNYU6544), and 21 months (CMNYU107, CMNYU6506). Viral RNA was extracted from patient plasma and the region from position 536 bp to 1200 bp (Hxb2 numbering engine) of each of the viral strains was amplified by nested polymerase chain reaction (PCR) (outer primers: 5'-GACCAGATCTGAGCCTGGGAGC-3' and 5'-TGAAAGCCTTTTCTTCTA-3'; inner primers: 5'-CAATAAA GCTTGCTTGGAG-3' and 5'-ATTTTGCATATAGGTAAT TT-3'), reverse transcribed into cDNA, and cloned, as described previously.⁶

Sequence alignment and phylogenetic analysis of the 664-bp sequence (536–1200 bp) by the neighbor-joining method (1000 bootstrap replicas), with genetic distances evaluated with the Kimura two-parameter method, were performed using MAFFT, Seaview, and MEGA4.^{11–14} A total of 13–24 clones per sample were sequenced and analyzed. The sequences have been deposited in GenBank with accession numbers JF791817–JF792044. Subjects CMNYU124/CMNYU129 and CMNYU6506/CMNYU6544 were found to be infected with genetically similar strains. These individuals may have been infected by the same virus as members of the same network of sexual partners. Importantly, repeat RNA extraction and PCR were performed separately for each sample using different plasma aliquots to ensure that these linkages are not a result of sample contamination or mislabeling.

Four study subjects were identified who were superinfected by subtype-discordant HIV-1 strains. The initial virus that infected patient CMNYU107 clustered with CRF01A1 in the region analyzed, while the superinfecting strain was classified as subtype G (Fig. 1). Eighteen clones from subject CMNYU107 were analyzed from time point 1 (month 0) and 24 clones from time point 2 (month 21). Within

FIG. 1. Phylogenetic analysis of 5' LTR-early gag (HxB2 location 526–1200 bp) quasi-species clones amplified from patient plasma samples. Most reference and study subject sequences have been omitted for clarity. Study sequences are identified by the subject number, study month number (in parentheses), and clone number. Reference sequences are highlighted by black circles.

the 35-nucleotide-long DIS sequence the adenine observed at position 722 of the initial virus was replaced with a guanine in the superinfecting viral strain, leading to a mismatch in the DIS stem. This change was observed in all the clones analyzed. Additionally, 33% of the initial virus clones showed a guanine at nucleotide position 717 while an adenine was observed in all other initial clones and all clones of the superinfecting virus (Fig. 2A). The nucleotide at location 717 flanks the palindromic loop sequence and is essential to the dimerization process by the formation of noncanonical interactions.¹⁵ Patient CMNYU124 was initially infected with a virus of subtype F2 (month 0) and superinfected with a CRF02_AG strain (month 9) (Fig. 1). Nineteen and 24 clones were analyzed of the initial and second virus strain, respectively. All clones from the initial virus contained an adenine at position

715 of the DIS sequence whereas 19 clones (79%) of the dually infecting virus had a guanine at this location, which destroys the self-complementarity of the loop. The remaining five clones of the dually infecting virus with an adenine at position 715 had position 717, the nucleotide next to the loop, changed from an adenine to a guanine (Fig. 2A). The viral subtypes detected in study subject CMNYU129 were classified as CRF02_AG and F2 at month 0 and month 9, respectively (Fig. 1). All 21 clones analyzed from time point 1 showed a guanine at position 715, leading to a mismatch in the palindromic loop region (data not shown). The final patient in this cohort superinfected with subtype-discordant strains was CMNYU6544. The study subject was initially infected with a subtype F2 (month 0) virus and then superinfected with CRF02_AG (month 12) (Fig. 1). Twenty-two clones were

FIG. 2. Dimerization initiation site (DIS) sequences of HIV-1 strains in superinfected patients. **(A)** DIS sequences (697–731bp, Hxb2 numbering engine) of two subjects superinfected with subtype-discordant viruses. Patients' sequences are identified by the subject number, study month number (in parentheses), and clone number. The arrows point to the single nucleotide changes between the two superinfecting viral strains. **(B)** DIS sequences (697–731bp, Hxb2 numbering engine) of two subjects superinfected with subtype-concordant viruses. Patients' sequences are identified by the subject number, study month number (in parentheses), and clone number.

examined of time point 1 and 23 clones of time point 2. Within the DIS sequence the adenine observed at position 717 of the initial virus was replaced with a guanine in all clones of the superinfecting viral strain. At position 710 an adenine was found in all clones of the initial virus compared to a guanine in the superinfecting virus. Both of these changes affect the nucleotides in the region flanking the loop. Furthermore, in 36% of the clones of the initially infecting virus the nucleotide adenine was exchanged with a guanine at position 719, which produces a mismatch in the DIS stem (data not shown).

Two study subjects infected with two HIV-1 strains of the same subtype were also analyzed and included CMNYU6506 and CMNYU6564. The pairs of concordant viruses were classified as subtype CRF02_AG in both patients (Fig. 1). The genetic distance in the analyzed region between the two time points in subject CMNYU6564 was determined as 5.2%. In subject CMNYU6506 the initial virus was isolated in month 3 (13 clones analyzed) and the superinfecting virus strain in month 24 (14 clones analyzed). In study subject CMNYU6564 the two virus strains were isolated from month 0 (15 clones analyzed) and month 3 (13 clones analyzed). All clones from the genetically concordant superinfecting strains revealed identical DIS sequences (Fig. 2B).

Previous *in vitro* studies showed that sequence identity among two viral strains is critical for dimerization and recombination^{7,8} and that dimerization of HIV-1 RNA sequences can be abolished by the introduction of single point mutations into the DIS loop.⁹ We observed a single nucleotide change within the DIS loop in two of the subtype-discordant superinfected study subjects, CMNYU124 and CMNYU129. In CMNYU124 79% of the analyzed clones had a replacement of adenine at position 715 with guanine and in CMNYU129 all clones had the mentioned mismatch that destroyed the self-complementarity of the loop. *In vitro* studies have shown that a guanine at this position greatly reduces dimerization. Additionally, a strong decrease of the thermal stability of dimers with a guanine at position 715 was observed.⁹

Most of the single nucleotide changes detected in the subtype-discordant HIV-1 superinfections were in the three bases flanking the self-complementary sequence in the loop. These nucleotides, which are most often adenines and almost

always purines, are highly conserved. While they are normally unpaired, it has been suggested that they stabilize the kissing loop dimer via the formation of noncanonical interactions.¹⁵ In three of the four study subjects (CMNYU107, CMNYU124, CMNYU6544) we found a guanine replacing the adenine at position 717 in one of the two superinfecting viruses in some or in all clones analyzed. While no *in vitro* experiments have analyzed the effect of the exact change from an adenine to a guanine, deletion of the flanking nucleotides was shown to affect the dimerization kinetics and the stability of the RNA dimers and can abolish dimerization.^{10,15}

Single nucleotide changes leading to a mismatch in the stem structure were observed in patients CMNYU107 and CMNYU6544 at positions 722 and 719, respectively. Clever *et al.* showed *in vitro* that mutations that disrupt base pairing in the stem abolish dimerization.¹⁰ Thus, the single nucleotide changes we observed in the DIS loop, the flanking nucleotides, and the stem region might prevent dimerization of the two superinfecting discordant HIV-1 strains and might be responsible for the lack of recombination in these study subjects *in vivo*.⁶ These data confirm that sequence identity plays an important role in HIV-1 recombination (Fig. 3).

Various factors such as viral replicative fitness, viral load, and host immune pressure are likely to affect the frequency at which two viral strains recombine *in vivo*. Additionally, viral features that increase reverse transcriptase pausing have been shown to promote recombination, as pausing contributes to template switching. These features include secondary RNA structures and RNA breakage.⁵ Recombinants between viruses with different DIS sequences have been identified in human populations; this suggests that other regions of the viral genome can contribute to dimer formation and recombination rate.^{7,16} Protein cofactors such as the nucleocapsid protein can also facilitate and stabilize RNA dimerization⁴ and thus have an effect on the recombination rate during virus replication.

In conclusion, we find that differences exist in the DIS of the subtype-discordant versus the subtype-concordant strains superinfecting HIV-1 patients, suggesting that dimerization and recombination are more likely to occur among concordant than among discordant superinfecting HIV-1 strains.

Position	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731
Patient ID																																			
107(0)	C	T	C	G	G	C	T	T	G	C	T	G	A	G	G	T	G	C	A	C	A/G	C	A	G	C	A	A	G	A	G	G	C	G	A	G
107(21)	C	T	C	G	G	C	T	T	G	C	T	G	A	G	G	T	G	C	A	C	A	C	A	G	C	G	A	G	A	G	G	C	G	A	G
124(0)	C	T	C	G	G	C	T	T	G	C	T	G	A	G	G	T	G	C	A	C	A	C	A	G	C	A	A	G	A	G	G	C	G	A	G
124(9)	C	T	C	G	G	C	T	T	G	C	T	G	A	G	G	T	G	C	A/G	C	A/G	C	A	G	C	A	A	G	A	G	G	C	G	A	G
129(0)	C	T	C	G	G	C	T	T	G	C	T	G	A	G	G	T	G	C	G	C	A	C	A	G	C	A	A	G	A	G	G	C	G	A	G
129(9)	C	T	C	G	G	C	T	T	G	C	T	G	A	G	G	T	G	C	A	C	A	C	A	G	C	A	A	G	A	G	G	C	G	A	G
6544(0)	C	T	C	G	G	C	T	T	G	C	T	G	A	A	G	T	G	C	A	C	A	C	A/G	G	C	A	A	G	A	G	G	C	G	A	G
6544(12)	C	T	C	G	G	C	T	T	G	C	T	G	A	G	G	T	G	C	A	C	G	C	A	G	C	A	A	G	A	G	G	C	G	A	G
6506(3)	C	T	C	G	G	C	T	T	G	C	T	G	A	G	G	T	G	C	A	C	G	C	A	G	C	A	A	G	A	G	G	C	G	A	G
6506(24)	C	T	C	G	G	C	T	T	G	C	T	G	A	G	G	T	G	C	A	C	G	C	A	G	C	A	A	G	A	G	G	C	G	A	G
6564(0)	C	T	C	G	G	C	T	T	G	C	T	G	A	G	G	T	G	C	A	C	A	C	A	G	C	A	A	G	A	G	G	C	G	A	G
6564(3)	C	T	C	G	G	C	T	T	G	C	T	G	A	G	G	T	G	C	A	C	A	C	A	G	C	A	A	G	A	G	G	C	G	A	G

FIG. 3. Summary of dimerization initiation site (DIS) sequences of HIV-1 superinfecting viruses in patients. The DIS sequence from nucleotide position 697–731 (Hxb2 numbering engine) of four patients superinfected with subtype-discordant (107, 124, 129, 6544) and two patients superinfected with subtype-concordant (6506, 6564) HIV-1 strains is displayed. Patients' sequences are identified by the subject number and study month number (in parentheses). Changes in nucleotides are highlighted in bold italic.

Acknowledgments

The authors are grateful to the individuals who have donated their blood for these studies. They wish to acknowledge the continued support of the Ministry of Public Health, Cameroon. Supported by Grant AI083142 from the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), National Cancer Institute (NCI) (CA153726), Fogarty International Center (TW001409), Center for AIDS Research (AI027742), and funds from the Department of Veterans Affairs (Merit Review Award and the Research Enhancement Program).

Author Disclosure Statement

No competing financial interests exist.

References

1. Korber B: HIV Molecular Immunology. Los Alamos National Laboratory.
2. Powell RL, Urbanski MM, Burda S, Kinge T, and Nyambi PN: High frequency of HIV-1 dual infections among HIV-positive individuals in Cameroon, West Central Africa. *J Acquir Immune Defic Syndr* 2009;50(1):84–92.
3. Hu WS and Temin HM: Genetic consequences of packaging two RNA genomes in one retroviral particle: Pseudodiploidy and high rate of genetic recombination. *Proc Natl Acad Sci USA* 1990;87(4):1556–1560.
4. Paillart JC, Shehu-Xhilaga M, Marquet R, and Mak J: Dimerization of retroviral RNA genomes: An inseparable pair. *Nat Rev Microbiol* 2004;2(6):461–472.
5. Onafuwa-Nuga A and Telesnitsky A: The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiol Mol Biol Rev* 2009;73(3):451–480, Table of Contents.
6. Powell RL, Lezeau L, Kinge T, and Nyambi PN: Longitudinal quasispecies analysis of viral variants in HIV type 1 dually infected individuals highlights the importance of sequence identity in viral recombination. *AIDS Res Hum Retroviruses* 2010;26(3):253–264.
7. Chin MP, Chen J, Nikolaitchik OA, and Hu WS: Molecular determinants of HIV-1 intersubtype recombination potential. *Virology* 2007;363(2):437–446.
8. Chin MP, Rhodes TD, Chen J, Fu W, and Hu WS: Identification of a major restriction in HIV-1 intersubtype - recombination. *Proc Natl Acad Sci USA* 2005;102(25):9002–9007.
9. Paillart JC, Marquet R, Skripkin E, Ehresmann B, and Ehresmann C: Mutational analysis of the bipartite dimer linkage structure of human immunodeficiency virus type 1 genomic RNA. *J Biol Chem* 1994;269(44):27486–27493.
10. Clever JL, Wong ML, and Parslow TG: Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA. *J Virol* 1996;70(9):5902–5908.
11. Galtier N, Gouy M, and Gautier C: SEAVIEW and PHYLO_WIN: Two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* 1996;12(6): 543–548.
12. Kumar S, Nei M, Dudley J, and Tamura K: MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 2008;9(4):299–306.
13. Katoh K and Toh H: Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 2008;9(4):286–298.
14. Kimura M: A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 1980;16(2):111–120.
15. Paillart JC, Westhof E, Ehresmann C, Ehresmann B, and Marquet R: Non-canonical interactions in a kissing loop complex: The dimerization initiation site of HIV-1 genomic RNA. *J Mol Biol* 1997;270(1):36–49.
16. Hill MK, Shehu-Xhilaga M, Campbell SM, Pombourios P, Crowe SM, and Mak J: The dimer initiation sequence stem-loop of human immunodeficiency virus type 1 is dispensable for viral replication in peripheral blood mononuclear cells. *J Virol* 2003;77(15):8329–8335.

Address correspondence to:

Phillipe Nyambi

Department of Pathology

New York University School of Medicine

c/o V.A. Medical Center

423 East 23rd Street, Room 18124N

New York, New York, 10010

E-mail: phillipe.nyambi@nyumc.org