Genetic Analysis of a UNAIDS HIV Type 1 from Brazil Revealed an Unexpected Recombination Pattern

Mario P.S. Chin and Natalia Ristic

EDITOR: The World Health Organization (WHO) Network for HIV Isolation and Characterization (now named UNAIDS) was established by the WHO Global Programme on AIDS to systematically isolate and characterize HIV strains from different parts of the world, and to obtain information and reagents that would facilitate HIV vaccine development.^{1,2} The laboratory network collected specimens from WHO-sponsored vaccine evaluation sites in Brazil, Rwanda, Thailand, and Uganda. This collection contains a significant number of intersubtype recombinants with diverse recombination patterns.^{3,4} Among the sequences of HIV-1 collected in Brazil by this network, we identified the sequences of a recombinant strain that have conflicting recombination patterns in bootscanning analysis.

The isolate in question is 92BR023, which was isolated from an asymptomatic heterosexual male from Proto Alegre, Brazil.² The 92BR023 isolate was genotyped previously, and six sequences derived from the viral genome have been reported.^{2–6} The sequence fragments included two overlapping regions in *gag* covering nucleotide positions 859–1587 and 1407–2131 (HXB2 numbering),^{3,5} two regions in *pol* spanning positions 2265–3440 and 4230–5064,^{4,6} and two overlapping sequences in *env* covering 7032–7310 and 7050–7400 (Fig. 1).⁷ Based on the genotyping data, it was shown that this isolate is a B/C intersubtype recombinant with a subtype C *gag*, a subtype C *pol*, and a subtype B *env*. When we analyzed these sequences of 92BR023 obtained from the Los Alamos HIV Sequence Database by bootscanning, however, a more complex recombination pattern was observed and one of the sequences appeared to generate conflicting results.

In the bootscanning analysis (SimPlot version 3.5.1),⁸ we used a subtype B and a subtype C HIV-1 from Brazil as references, and a subtype A and a subtype K virus from other regions as the outgroup.⁹ Comparing the two sequences within the *gag* of 92BR023, the fragment covering the matrix (MA) and capsid (CA) genes belong to subtype C, whereas the part consisting of CA and nucleocapsid (NC) is subtype B (Fig. 1). This is not consistent with previous reports indicating that 92BR023 has a subtype C *gag*. It is possible that a recombination breakpoint is present in the *gag* gene, but out of the 181-nt overlapping sequences of the two fragments, there were 12 mismatches. The genetic distance of 0.066 (12/181) is similar to the genetic distance between subtypes B and C *gag*, suggesting that there might be errors in the HIV-1 sequences.

We then analyzed the *pol* and *env* of 92BR023 and identified additional recombination breakpoints. Two recombination breakpoints were identified in the bootscanning analysis of the regions covering protease (PR) and reverse transcriptase (RT) genes (Fig. 1). These breakpoints resulted in a short subtype B sequence within the subtype C *pol*. One potential breakpoint was observed in the integrase (IN) gene. In addition, two recombination breakpoints were identified in the short sequence covering C2–C3 of gp120. None of these recombination breakpoints in *pol* and *env* has been reported previously. The results indicate that 92BR023 appears to be a more complex B/C recombinant than the one originally suggested.^{3,4} Here, we determined the sequence of the full-length genome of 92BR023 to map out the recombination patterns of the virus and to confirm the accuracy of the GenBank entries associated with the isolate.

We isolated RNA directly from the virus stock obtained from the AIDS Reagent Program and converted the RNA to cDNA using SuperScript III reverse transcriptase (Invitrogen). The cDNA was amplified using the FastStart High Fidelity PCR System (Roche) in four overlapping fragments covering the full-length genome of 92BR023. The PCR products were sequenced with overlapping primers, and the resulting sequence contigs were assembled with the Staden Package (PCR and sequencing primer sequences are available on request).¹⁰ Every nucleotide was identified by at least two sequence contigs to ensure the accuracy of the DNA sequence. The assembled viral sequences were aligned with the reference and outgroup sequences using Clustal X (version 1.8.3).¹¹

Bootscanning was carried out for the 92BR023 sequences with the same references and outgroups described above. We found that the 92BR023 is indeed a B/C recombinant, but the recombination pattern is quite complex (Fig. 1). Notably, the *gag* of the recombinant is subtype C and does not seem to have recombined with another subtype. This finding ruled out the possibility that a recombination breakpoint is present in *gag* and illustrated that the GenBank entry, AY090758, does not match the sequences of the 92BR023. We compared AY090758 with the same region from the full-length 92BR023 sequence and found that the sequences shared only 89.5% similarity, thus confirming that the GenBank entry is incorrect (Table 1).

In the *pol*, we identified two short subtype B regions within the subtype C sequence (Fig. 1). The recombination breakpoints in the RT genes are consistent with the bootscanning

FIG. 1. Bootscanning of the 92BR023 viral genome reveals numerous recombination breakpoints. Five GenBank entries associated with the 92BR023 HIV-1 (accession numbers were shown on top of each analysis) were compared with a subtype A (AY521631), a subtype B (AY173956), a subtype C (AY727522), and a subtype K (AJ249239) HIV-1 sequence in bootscanning analyses. The results are shown as plots of percentage of permuted trees against nucleotide positions above the HIV-1 genomic structure. A window/step size of 200/10 was used to analyze the sequences within the *gag* or *pol* region, whereas 100/5 was used to examine the *env* sequence. Black bars indicated by arrows represent the coverage of each sequence on the viral genome. The genomic sequence of 92BR023 (nucleotide positions 818–9459) was compared with the same references described above and the bootscanning plot is shown at the bottom. MA, matrix; CA, capsid; NC, nucleocapsid; PR, protease; RT, reverse transcriptase; IN, integrase. GenBank accession number of 92BR023 full-length genomic sequence: HM100716. (Color image can be found at www.liebertonline.com/aid).

result from the GenBank entry AF009376. We also demonstrated the presence of a recombination breakpoint at the start of the IN gene, which confirms the bootscanning analysis of AF458237. Two separate studies indicated that 92BR023 has a subtype B *env*.^{2,5} Both studies were based on the results from genotyping the short V3 region and phylogenetic analysis using the C2–V3 or C2–C3 region of the *env*. Our analysis showed that 92BR023 has a subtype B V3, but it also has several recombination breakpoints in the *env* (Fig. 1). Neither study identified the observed recombination in the C2–C3 region in the phylogenetic analyses, probably because phy-

logenetic inference is not a sensitive method to detect recombination, especially when the region of interest is relatively short (300–400 bp).

Finally, we verified the accuracy of the GenBank entries associated with 92BR023 by comparing the entries with the corresponding regions of the full-length sequence. Except for AY090758, which was described previously, all five sequences shared a high degree of similarity with the full-length sequence of 92BR023 (Table 1), with one to five nucleotide mismatches probably generated from PCR and sequencing errors or due to the fact that our sequence was obtained from

TABLE 1. NUCLEOTIDE SIMILARITY BETWEEN 92BR023 AND GENBANK ENTRIES

Query sequence ^a	Region ^b	Nucleotide length	Match number	Similarity (%)
U86559	MA/CA	710	705	99.3
AY090758	CA/NC	731	654	89.5
AF009376	PR/RT	1176	1175	99.9
AF458237	IN	835	831	99.5
U08709	Env	273	272	99.6
U08779	Env	345	342	99.1

^aQuery sequences were downloaded from the Los Alamos HIV Sequence Database. Accession numbers are shown.

^bHIV-1 genes covered by the query sequences. MA, matrix; CA, capsid; NC, nucleocapsid; PR, protease; RT, reverse transcriptase; IN, integrase; Env, envelope.

virus passage different from those in previous studies. Altogether, we presented the full-length sequence of 92BR023 and mapped the recombination pattern of the virus. We also showed that one of the GenBank entries associated with 92BR023 does not match the sequence of the virus. The full-length sequence of 92BR023 highlights the complexity of HIV-1 B/C intersubtype recombinants in Brazil and will help to elucidate the biological and antigenic properties of HIV-1 B/C recombinants.

Acknowledgments

HIV-1 92BR023 (catalog no. 1782) was obtained from The UNAIDS Network for HIV Isolation and Characterization through the AIDS Research and Reference Reagent Program. This work was supported by internal funds of the Aaron Diamond AIDS Research Center and National Institutes of Health Grant DA026293.

Author Disclosure Statement

No competing financial interests exist.

References

1. Osmanov S, Heyward WL, and Esparza J: The World Health Organization Network for HIV Isolation and Characterization: Summary of a pilot study. *AIDS Res Hum Retroviruses* 1994;10(11):1325–1326.
2. WHO Global Programme on AIDS: HIV type 1 variation in World Health Organization-sponsored vaccine evaluation sites: Genetic screening, sequence analysis, and preliminary biological characterization of selected viral strains. WHO Network for HIV Isolation and Characterization. *AIDS Res Hum Retroviruses* 1994;10(11):1327–1343.
3. Cornelissen M, Kampinga G, Zorgdrager F, and Goudsmit J: Human immunodeficiency virus type 1 subtypes defined by env show high frequency of recombinant gag genes. The UNAIDS Network for HIV Isolation and Characterization. *J Virol* 1996;70(11):8209–8212.
4. Burns CC, Gleason LM, Mozaffarian A, Giachetti C, Carr JK, and Overbaugh J: Sequence variability of the integrase protein from a diverse collection of HIV type 1 isolates representing several subtypes. *AIDS Res Hum Retroviruses* 2002;18(14):1031–1041.
5. Ball SC, Abrahams A, Collins KR, *et al.*: Comparing the *ex vivo* fitness of CCR5-tropic human immunodeficiency virus type 1 isolates of subtypes B and C. *J Virol* 2003;77(2):1021–1038.
6. Cornelissen M, van den Burg R, Zorgdrager F, Lukashov V, and Goudsmit J: pol gene diversity of five human immunodeficiency virus type 1 subtypes: Evidence for naturally occurring mutations that contribute to drug resistance, limited recombination patterns, and common ancestry for subtypes B and D. *J Virol* 1997;71(9):6348–6358.
7. WHO Network for HIV Isolation and Characterization: HIV type 1 variation in World Health Organization-sponsored vaccine evaluation sites: Genetic screening, sequence analysis, and preliminary biological characterization of selected viral strains. *AIDS Res Hum Retroviruses*. 1994;10(11):1327–1343.
8. Lole KS, Bollinger RC, Paranjape RS, *et al.*: Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 1999;73(1):152–160.
9. Moore MD, Chin MPS, and Hu WS: HIV-1 recombination: An experimental assay and a phylogenetic approach. In: *HIV Protocols*, Vol. 485, 2nd ed. (Prasad V, Kalpana G, Eds.). Humana Press, Inc., New York, 2009, pp. 87–105.
10. Staden R: The Staden sequence analysis package. *Mol Biotechnol* 1996;5(3):233–241.
11. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, and Higgins DG: The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997;25(24):4876–4882.

Address correspondence to:

Mario P.S. Chin
Aaron Diamond AIDS Research Center
The Rockefeller University
455 First Avenue
New York, New York 10016

E-mail: mchin@adarc.org