Reverse Transcriptase and Protease Sequence Evolution in Two HIV-1–Infected Couples

*Sarah Palmer, †Dominique Vuitton, ‡Matthew J. Gonzales, †Agnès Bassignot, and ‡Robert W. Shafer

*HIV Drug Resistance Program, National Cancer Institute, National Institutes of Health, Frederick, Maryland, U.S.A.; †Université de Franche-Comté, Centre Hospitalier Universitaire, Besançon, France; and ‡Division of Infectious Diseases and Geographic Medicine, Stanford University, Stanford, California, U.S.A.

Summary: We analyzed the reverse transcriptase (RT) and protease sequences of HIV-1 isolates obtained over 7 years from two couples with known transmission histories. Phylogenetic trees constructed from the sequence data reflected the known transmission histories, despite the fact that the drug resistance mutations were most consistent with the drug treatment histories. However, the RT sequences from one couple diverged by 2.9% even before therapy was begun, and three (0.9%) of 339 unrelated individuals had viruses that shared a common ancestor with sequences from the recipient member of the couple but not with sequences from the transmitter. The divergence between the first two isolates from this couple is consistent with a pre-transmission interval during which the transmitter developed a heterogeneous virus population. The closeness between the three controls and the recipient’s first RT sequence may indicate slower evolution on the branches of the control sequences. Although the RT and protease genes contain phylogenetic information, they are sub-optimal for reconstructing transmission history because the genetic distance between RT and protease isolates from unrelated individuals may occasionally approximate the distance between RT and protease isolates from related individuals. Key Words: HIV—Antiretroviral resistance—HIV transmission history—AF487122–AF487139.

The error-prone replication of HIV-1 leads to the development of a viral quasispecies within individuals and to virus diversification at the population level as the virus spreads between infected individuals (1). Because the genetic diversity between individuals exceeds the diversity within individuals, it has been possible—using env and parts of gag—to perform molecular epidemiological studies of HIV transmission (2,3). Indeed, recent epidemiological studies and legal cases concerning the transmission of HIV between individuals have relied for the most part on env sequences and to a lesser degree on gag sequences (4,5).

The reverse transcriptase (RT) and protease genes are frequently sequenced in clinical settings to assist physicians in selecting antiretroviral therapy (6), but due to their conserved nature, these genes have not been examined in a phylogenetic context. Not only is there less variability in the RT and protease genes, but also the variability that is present is often caused by drug therapy rather than epidemiological factors. Nonetheless, recent studies have shown that RT and protease display enough interindividual variation to make subtype determination possible (7–9). In addition, interindividual variation is used to exclude PCR contamination since it is unlikely for the RT and protease sequences from two patients to be <2% different (10).

Because the utility of RT and protease for tracking transmission has not been examined, we analyzed the RT and protease sequences of HIV-1 isolates from two
<table>
<thead>
<tr>
<th>NRTI</th>
<th>NNRTI</th>
<th>PI</th>
<th>Reverse transcriptase</th>
<th>Protease</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>37 41 49 64 65 69 70</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>75 83 89 118 177 184</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>210 211 213 214 215</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>214 215 217 218 219</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>244 245 246 247 248</td>
<td></td>
</tr>
</tbody>
</table>

Patient A

Jan 92 — — — K R K K P V
Apr 96 AZT, DDC — — L K K Y D N K P V
Mar 97 AZT, 3TC — — L K E V W K Y D S K P V
Sep 97 AZT, 3TC — RTV L K E V W K Y D S R P V I
Jan 92 — — — R K I D M K P V
Oct 96 AZT, DDC, 3TC — — L R N R S I V L Q R D K P V
Jun 97 D4T, 3TC — IDV L R T N R S I V L E Q D I K V P V
May 98 ABC EFV IDV, NFV, IDV, RTV/SQV R H G N R S I L E Q F I D M N M V K P V A
Sep 98 ABC EFV RTV/SQV L R H G N R S I L E Q D I D V K V P V A

Beneath the headings Reverse transcriptase and Protease are the positions that differed from consensus in at least one of the sequences from patients A and B. The letter directly beneath the position is the consensus amino acid at the position. The letters beneath that are the mutations present in the patient sequences. The presence of more than one mutation indicates a mixture.

NRTI, nucleoside RT inhibitors; NNRTI, nonnucleoside RT inhibitors; PI, protease inhibitors; AZT, zidovudine; DDC, zalcitabine; 3TC, lamivudine; D4T, stavudine; ABC, abacavir; EFV, efavirenz; RTV, ritonavir; IDV, indinavir; NFV, nelfinavir; SQV, saquinavir.

TABLE 1. Antiretroviral treatments and reverse transcriptase (RT) and protease mutations of an HIV-1–infected couple (patients A and B)
couples with known transmission histories to ascertain the extent to which the phylogenetic relatedness of these genes was preserved in patients receiving antiretroviral therapy.

METHODS

Patients

Plasma samples were obtained from two HIV-1–infected couples attending an outpatient clinic (Department of Internal Medicine and Clinical Immunology, University Hospital, Besançon, France) between 1991 and 1998. The HIV-1 transmission history within each couple was known. Patient A, a male intravenous drug user (intravenous drug use between 1978 and 1982), was found to be HIV-1–seropositive in November 1991 when his female sexual partner (patient B), who had no HIV-1 risk factors other than contact with patient A, was found to be HIV-1–seropositive. Patient A also was infected with HCV and had symptoms of chronic active hepatitis; patient B remained HCV-negative for the whole study period. Patient D was infected with HIV-1 in 1987; his male partner (patient E) seroconverted in August 1991. Both recipients (patients B and E) had symptoms of primary HIV

<table>
<thead>
<tr>
<th>NRTI</th>
<th>NNRTI</th>
<th>PI</th>
<th>RT</th>
<th>Protease</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>June 91</td>
<td>——</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>May 95</td>
<td>AZT, DDI, DDC, 3TC</td>
<td>—</td>
<td>—</td>
<td>M</td>
</tr>
<tr>
<td>Jul 96</td>
<td>AZT, 3TC</td>
<td>—</td>
<td>—</td>
<td>SQV</td>
</tr>
<tr>
<td>Sep 98</td>
<td>D4T, DDI, AZT, 3TC</td>
<td>NVP, EFV</td>
<td>SQV, IDV, NFV, RTV</td>
<td>M</td>
</tr>
<tr>
<td>Patient E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dec 91</td>
<td>——</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Jun 96</td>
<td>AZT, DDI, DDC</td>
<td>—</td>
<td>—</td>
<td>N</td>
</tr>
<tr>
<td>Feb 97</td>
<td>D4T, 3TC</td>
<td>—</td>
<td>—</td>
<td>L</td>
</tr>
<tr>
<td>Sep 98</td>
<td>D4T, 3TC</td>
<td>—</td>
<td>—</td>
<td>IDV</td>
</tr>
</tbody>
</table>

**TABLE 2.** Antiretroviral treatments and reverse transcriptase (RT) and protease mutations of an HIV-1–infected couple (patients D and E)
infection, and antigen to HIV-1 was found in their sera within the 45 days preceding seroconversion. Patient B was seronegative in 1988 (volunteer screening, according to the French National Agency for AIDS Research) when she gave birth to a seronegative child; her partner (patient A) was not tested for HIV at that time.

**Sequence Analysis of Patient Strains**

A 1.3-kb fragment of cDNA encompassing HIV-1 protease and the first 300 codons of RT was sequenced from patient plasma as previously described (11–13). Sequences were compared with the consensus B sequence to derive a list of mutations (14). Sequences from the two HIV-1–infected couples and control patients (not described previously) were submitted to the GenBank database (AF487122–AF487139).

**Phylogenetic Analysis**

Phylogenetic trees were constructed from the complete protease sequence, the 5'-polymerase coding region of the RT (codons 1–250), and the concatenation of the protease and RT sequences by means of fastDNAmL (15) and Phylogenetic Analysis Using Parsimony (PAUP*) (16). Neighbor-joining maximum parsimony and maximum likelihood trees were created using a variety of nucleotide substitution models including the uncorrected, Kimura 2 parameter model (which accounts for a transition/transversion bias) and the HKY85 model with gamma distribution (HKY85 + I), which accounts for a transition/transversion bias, variable base frequency data, and variable substitution rates at different nucleotide positions) (15,16).

Subtype B sequences from the HIV Outpatient Clinic at Université de Franche-Comté (Besançon, France) and the Stanford University HIV RT and Pro tease Sequence Database were included in the trees as

![Neighbor-joining tree constructed from a combination of the reverse transcriptase and protease genes using the HKY85 nucleotide substitution model with a gamma distribution (HKY85 + I). Gray boxes highlight sequences from the HIV-1–infected couples. Bootstr ap values are provided for the most recent common ancestor of each transmission pair and of the more recent sequences from each patient within the pair.](image)
controls (17). To minimize bias introduced by HIV drug therapy, trees were also created following the removal of nucleotides at positions associated with drug resistance (protease codons 10, 20, 24, 30, 32, 36, 46, 47, 48, 50, 53, 54, 63, 71, 73, 74, 77, 82, 84, 90, and 93; RT codons 41, 62, 65, 67, 69, 70, 74, 75, 77, 98, 100, 101, 103, 106, 108, 115, 116, 151, 179, 181, 184, 188, 190, 210, 215, 219, 225, and 236).

RESULTS

Plasma samples collected from each patient at four or five time points between 1991 and 1998 were genotyped. The antiretroviral treatment histories and RT and protease mutations are shown in Tables 1 and 2.

Figure 1 shows the neighbor-joining tree constructed from combined protease and RT sequences using the HKY85 + F nucleotide substitution model. Bootstrap values are provided for the most recent common ancestor (MRCA) of each transmission pair and the more recent sequences from each patient within the pair. Maximum likelihood and maximum parsimony trees constructed from complete sequences and sequences lacking positions associated with the development of drug resistance had branching patterns identical to those of the neighbor-joining tree.

The sequences obtained from patient D in July 1991 were closest to the MRCA for the sequences from patients D and E, consistent with the history that the virus was transmitted from patient D to patient E at about that time. The bootstrap value at the MRCA node of patients D and E was 98.

The sequence obtained from patient A in January 1992 was closest to the MRCA for the sequences from patients A and B, consistent with the transmission of virus from patient A to patient B at about that time. However, the bootstrap value at the MRCA node for patients A and B was only 42. The genetic distance between the sequences from patients A and B at their initial time points was also higher than expected (2.9%, RT; 0.35%, protease).

Because of the high genetic distance between the initial RT sequences from patients A and B and because of the low bootstrap value at the their MRCA node, we constructed neighbor-joining, maximum likelihood, and maximum parsimony trees using RT sequences from the two patients and an additional 316 subtype B GenBank sequences for a total 339 control sequences (including the 23 initial control sequences). In these trees, one of the original control sequences...
DISCUSSION

We examined the phylogenetic relatedness of RT and protease sequences of HIV-1 isolates from two HIV-1–infected French couples with known transmission histories. For both couples, a precise date for the recipients’ infections could be assessed because of clinical and biologic signs of primary HIV-1 infection. The phylogenetic relatedness of these virus isolates was examined over time even as the patients were receiving different antiretroviral treatment regimens. The phylogenetic trees reflected the known transmission histories, despite the fact that the drug resistance mutations were most consistent with the patients’ drug treatment histories.

However, the initial RT sequences from one couple were relatively divergent. Indeed, three control sequences from unrelated individuals were found to share an MRCA node with recipient sequences (patient B) but not with transmitter sequences (patient A). The divergence between the first two time points for patients A and B is most consistent with a long pretransmission interval during which time patient A’s virus developed into a heterogeneous virus population (18). Patient A was presumably infected between 1978 and 1982; patient B was his sexual partner between the end of 1982 and 1991. She was found to be HIV-1–seronegative in 1988 and had clinical and biologic signs of HIV-1 infection in September 1991. The closeness between the three control sequences and patient B’s first sample may represent slower evolution in the RT along the branches of the control sequences possibly because of the requirements for enzymatic function.

This study suggests that although the RT and protease genes contain phylogenetic information, they are suboptimal for reconstructing transmission history because the genetic distance between RT and protease isolates obtained from unrelated individuals may occasionally approximate the distance between RT and protease isolates from related individuals. Sequencing of other more variable HIV-1 genes will continue to be required for molecular epidemiological and forensic analyses.

Acknowledgment: The authors thank Victoria Hellmann for excellent technical assistance.

REFERENCES


JAIDS Journal of Acquired Immune Deficiency Syndromes, Vol. 31, No. 3, November 1, 2002