Frequent polymorphism at drug resistance sites in HIV-1 protease and reverse transcriptase

Mary Kearney\textsuperscript{a}, Sarah Palmer\textsuperscript{a}, Frank Maldarelli\textsuperscript{a}, Wei Shao\textsuperscript{b}, Michael A. Polis\textsuperscript{c}, JoAnn Mican\textsuperscript{c}, Diane Rock-Kress\textsuperscript{d}, Joseph B. Margolick\textsuperscript{e}, John M. Coffin\textsuperscript{a} and John W. Mellors\textsuperscript{f}

\textbf{Background:} Failure of antiretroviral therapy may result from the selection of pre-existing, drug-resistant HIV-1 variants, but the frequency and type of such variants have not been defined.

\textbf{Objective:} We used single genome sequencing (SGS) to characterize the frequency of polymorphism at drug resistance sites in protease (PR) and reverse transcriptase (RT) in plasma samples from antiretroviral naive individuals.

\textbf{Methods:} A total of 2229 \textit{pro-pol} sequences in 79 plasma samples from 30 patients were analyzed by SGS. A mean of 28 single genome sequences was obtained from each sample. The frequency of mutations at all PR and RT sites was compared to those associated with drug resistance.

\textbf{Results:} We detected polymorphism at one or more drug resistance sites in 27 of 30 (90\%) patients. Polymorphism at positions 179 and 215 of RT was most common, both occurring in 23\% of patients. Most (68\%) of other drug resistance sites were polymorphic with an average of 3.2\% of genomes per sample containing at least one variant from wild type. Seven drug resistance sites were polymorphic in more than 1\% of genomes: PR position 33; RT positions 69, 98, 118, 179, 210, and 215. Although frequencies of synonymous polymorphism were similar at resistance and nonresistance sites, nonsynonymous polymorphism were significantly less common at drug resistance sites, implying stronger purifying selection at these positions.

\textbf{Conclusions:} HIV-1 variants that are polymorphic at drug resistance sites pre-exist frequently as minor species in antiretroviral naive individuals. Standard genotype techniques have grossly underestimated their frequency.

\textcopyright{} 2008 Wolters Kluwer Health | Lippincott Williams & Wilkins

\textit{AIDS} 2008, \textbf{22}:497–501
Introduction

The emergence of resistance to antiretroviral drugs is a major obstacle to their long-term efficacy. Minor populations of drug-resistant variants present before therapy have been hypothesized to be an important cause of treatment failure [1,2]. To date, the frequency and type of such mutants have not been well defined in vivo. Theoretical considerations predict that the frequency of variants in a large replicating virus population will be determined by a number of factors, including their mutation rate and selection coefficient, and the size of the replicating population [3]. Comparison of the frequencies of nonsynonymous changes at drug resistance sites should therefore provide insight into the relative selective forces acting on the corresponding positions in protease (PR) and reverse transcriptase (RT). In the present study, we used single-genome sequencing (SGS) to assess the frequency, diversity and linkage of mutations at drug resistance sites in antiretroviral naive, HIV-1-infected individuals, compared to the frequency of mutations at nondonor drug resistance sites [4].

Materials and methods

Plasma samples were obtained from 30 patients attending the National Institute of Allergy and Infectious Diseases (NIAID) Critical Care Medical Department of National Institutes of Health (NIH), Bethesda, Maryland, or enrolled in the AIEDRP study sponsored by NIAID and obtained from Joe Margolick at Johns Hopkins University, Baltimore, Maryland. The study was approved by the institutional review board of NIAID, NIH, Bethesda, Maryland, and all individuals provided written informed consent. Additional patient samples with evidence of drug-resistant variants, pro-pol diversity, single genome sequencing (SGS), treatment-naive patients.

Viral RNA from plasma containing 5000–10 000 copies of HIV-1 RNA was extracted and used for cDNA synthesis as previously described [4,5]. To obtain PCR products for SGS, the cDNA was diluted until approximately 30% of the PCR reactions yielded DNA product. cDNA was added to the PCR mix containing 2 mmol/l MgCl₂, 0.04 units Invitrogen Platinum Taq HiFi and 200 nmol/l primers: 1849+ (5’- GATGACAGCATGTGAC-3’), and 3500- (5’- CAGTTAGTGGTATTACTTCTGT-3’) producing a 1.5 kb amplicon containing the p6 region of gag, pro and the first 950 nt of pol. Positive PCR reactions were sequenced. The quality sequences from each plasma sample were aligned and compared to the HIV-1 subtype B consensus sequence using Clustal W software [6,7]. To perform standard ‘population’ genotype analyses, cDNA was generated as described above but was not diluted before PCR amplification and sequencing.

All sequences were analyzed for polymorphism at sites associated with drug resistance using the HIVseq Program available through the Stanford University HIV Drug Resistance Database (class I-IV definitions). Polymorphism was defined as any change from the subtype B consensus sequence (http://hivdb.stanford.edu). Frequency estimates were normalized to account for variations in the number of sequences obtained per plasma sample. We obtained a mean of 28 genomes per patient sample. Analyses of synonymous versus nonsynonymous mutations were performed using a software program written by Gary Smythers, PhD of Science Applications International Corporation (SAIC), Frederick, Maryland (details available upon request). Measurements of genetic diversity were calculated with a software program based on the average pairwise distance model. The duration of infection was determined in 20 of 31 patients by the onset of symptoms and a nonreactive enzyme-linked immunosorbent assay or negative western blot with detectable HIV-1 RNA. Longitudinal samples were available for up to 5 years from recently infected patients. The minimum duration of infection was used for chronically infected patients and was determined by the date of diagnosis and the patient exposure history.

Results

A total of 2229 sequences was obtained and analyzed from 79 plasma samples from 30 patients. Twenty-nine of the patients were male. The median [interquartile range (IQR)] age was 35 (29–38), CD4 cell count was 415 cells/μl (298–548), HIV-1 RNA was 46 490 copies/ml (16 230–204 170), and duration of infection was 183 days (71–592). First, we calculated the frequency of polymorphism at drug resistance sites and nonresistance sites. In PR, the frequency of nonsynonymous polymorphism at drug resistance sites was 32-fold lower than at nonresistance sites. In RT, the frequency of nonsynonymous polymorphism at drug resistance sites was 32-fold lower than at nonresistance sites. In PR, the frequency of nonsynonymous polymorphism at drug resistance sites was 32-fold lower than at nonresistance sites. In RT, the frequency of nonsynonymous polymorphism at drug resistance sites was 32-fold lower than at nonresistance sites. The differences in the frequency of polymorphism between resistance and nonresistance sites were statistically significant (P = 6 × 10⁻²⁸ in PR; P = 0.01 in RT). Highly polymorphic nonresistance sites (≥ 25% frequency) were
codons 12, 35, 36, 37, 41, 62, 63, 77 and 93 in PR and codons 35, 83, 122, 123, 135, 200, and 211 in RT.

To characterize HIV-1 polymorphism at sites associated with drug resistance, we plotted the frequency of nonsynonymous changes at each drug resistance site by drug class (Fig. 1). Polymorphism at drug resistance sites was more prevalent in RT (23 polymorphic of 31 total sites) than in PR (nine of 17). However, most of the variant positions in PR contained changes known to confer drug resistance (six of nine). In RT, fewer of the substitutions were known to confer drug resistance (10 of 23). Overall,
sequences in 45% of patients were polymorphic at one or more sites associated with PI resistance, 68% at one or more sites associated with nucleoside reverse transcriptase inhibitor (NRTI) resistance, and 66% at one or more sites associated with non-nucleoside reverse transcriptase inhibitors (NNRTI) resistance. Positions 179 and 215 in RT were most frequently polymorphic, both occurring in 23% of patients (Fig. 1a). In total, 90% of the anti-retroviral naive patients studied had at least one non-synonymous change from wild-type at a drug resistance site, and 56% had at least one change known to confer drug resistance.

We also determined the percentage of viral genomes polymorphic at drug resistance sites, averaged over all genomes, and plotted them by drug class (Fig. 1b). The cumulative frequency of nonsynonymous polymorphism at drug resistance sites was 0.2% in PR and 1.5% in RT. Of these, 0.03% in PR and 0.61% in RT were mutations known to confer drug resistance. Overall, known drug resistance mutations were detected in 350 of the 2229 genomes. Standard population genotype analyses identified polymorphism in only a small subset (one in PR and five in RT; asterisks in Fig. 1) of the resistance sites found to be polymorphic by SGS.

Linkage of polymorphism at drug resistance sites was found on genomes in eight of 30 patients (data not shown). Of these eight, only two patients showed linkage of known drug-resistance mutations. In the first patient, a G190E mutation was linked to D30N and M46I mutations in protease. In the second patient, M41I in RT was linked on the same genome with T69N and V179E. In both cases linked mutations were found as minority variants in the population.

Longitudinal plasma samples, obtained from two to seven time points over a span of up to 7 years, were used to analyze variation in polymorphism in patients over time (Fig. 2). All patient samples were categorized into one of three groups based on the estimated time since infection. The first group consisted of samples collected within 1 year of infection, the second group, samples collected within 1–5 years after infection, and the third group, samples collected 5–15 years after infection. The diversity of each sample was measured for all sites in pre-pol as well as for drug resistance sites only. Both the total diversity and the diversity at drug resistance sites increased with greater time post-infection, maintaining approximately the same ratio (0.79) across all groups. Thus, the frequency of polymorphism at drug resistance sites increased with the time after infection and with the overall viral diversity.

**Discussion**

We used SGS to characterize the frequency and types of polymorphism at drug-resistance sites in antiretroviral naive, HIV-1-infected individuals. Our data show that polymorphism at drug resistant sites is common and is grossly underestimated by standard genotyping methods. Because much of the polymorphism detected by SGS was present in only one or a few genomes in a sample, it was below the detection limit of standard population genotyping.

Although polymorphism at drug resistance sites was common, these sites were significantly more conserved than nonresistance sites. The greater conservation of resistance sites probably reflects a stronger selective disadvantage of any mutation and hence, the lower frequency of polymorphism, most likely relating to their location near structurally and functionally constrained regions of PR and RT. This finding supports the results presented in earlier reports of conservation of resistance sites relative to nonresistance sites using bulk sequencing techniques or cloning [8–11].

The extent of polymorphism at drug resistance sites varied by gene and by drug class. The NNRTI resistance associated sites were most frequently polymorphic (1.8% of total genomes), followed by NRTI (1.2%), and by PI (0.2%). The higher frequency of polymorphism at drug resistance sites in RT may be due to lesser constraints on these sites in the enzyme compared to PR. Interestingly, although polymorphism in PR was found to be less frequent than in RT, mutations detected in PR were more often known drug resistance mutations (Table 1).

Polymorphism at drug resistance sites increased with time and with the total genetic diversity of the virus population. Patient samples collected > 5 years post-infection contained highly diverse virus populations and had more frequent polymorphism at drug resistance sites. This data suggests that the longer patients are infected with HIV, the higher their viral diversity will be, and the greater likelihood they will have virus carrying resistance.
mutations. The clinical significance of this observation remains to be determined.

In summary, this study reveals that polymorphism at sites associated with drug resistance exists frequently prior to antiretroviral therapy, and that its frequency is underestimated by standard, population genotyping methods. Although pre-existing drug resistance mutations were often detected as minor variants, very few instances of linked drug resistance mutations were found. The absence of such linkage is consistent with the efficacy of combination antiretroviral therapy. The impact of these nonlinked, pre-existing, drug resistance mutations on antiretroviral therapy remains to be determined.

Acknowledgement

We thank Dr Robert Stephens and Dr Gary Smythers of SAIC, Frederick for aid with sequence analyses, Linda Apuzzo of Johns Hopkins University for aid in obtaining clinical specimens and sample information, Christopher Kearney of NCI-Frederick for aid with many helpful conversations. We also express our gratitude to the patients for participating in these studies. This work was supported in part by a grant (AI 41532) to J.B.M. and a NCI contract to J.W.M. (SAIC 20XS190A). This work was done in partial fulfillment of thesis requirements for M.K. This research was supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research.

Table 1. Polymorphism at drug resistance sites detected by single genome sequencing (SGS) in HIV infection.

<table>
<thead>
<tr>
<th>Position in PR</th>
<th>PI mutations* (# genomes)b</th>
<th>Position in RT</th>
<th>NRTI mutations* (# genomes)b</th>
<th>Position in RT</th>
<th>NNRTI mutations* (# genomes)b</th>
</tr>
</thead>
<tbody>
<tr>
<td>L23</td>
<td>none</td>
<td>M41</td>
<td>I(3)</td>
<td>A98</td>
<td>S(124)</td>
</tr>
<tr>
<td>L24</td>
<td>none</td>
<td>E44</td>
<td>K(3)</td>
<td>LT00</td>
<td>F(11)</td>
</tr>
<tr>
<td>D30</td>
<td>N(3) G(1)</td>
<td>A62</td>
<td>T(2)</td>
<td>K101</td>
<td>R(12)</td>
</tr>
<tr>
<td>V32</td>
<td>none</td>
<td>K65</td>
<td>none</td>
<td>K103</td>
<td>R(12)</td>
</tr>
<tr>
<td>L33</td>
<td>V(149)</td>
<td>D67</td>
<td>N(3) G(2) A(1) R(1)</td>
<td>V106</td>
<td>A(3) L(1) L(1)</td>
</tr>
<tr>
<td>M46</td>
<td>I(8)</td>
<td>T69</td>
<td>N(26) S(27) P(3)</td>
<td>V108</td>
<td>K(1) L(1)</td>
</tr>
<tr>
<td>I47</td>
<td>none</td>
<td>K70</td>
<td>E(1)</td>
<td>V179</td>
<td>D(11) E(86) I(36) A(1)</td>
</tr>
<tr>
<td>G48</td>
<td>R(1)</td>
<td>L74</td>
<td>V(1)</td>
<td>Y181</td>
<td>none</td>
</tr>
<tr>
<td>I50</td>
<td>none</td>
<td>V75</td>
<td>none</td>
<td>Y188</td>
<td>none</td>
</tr>
<tr>
<td>F53</td>
<td>L(2)</td>
<td>F77</td>
<td>none</td>
<td>G190</td>
<td>E(2) A(1) R(1)</td>
</tr>
<tr>
<td>L54</td>
<td>V(1)</td>
<td>Y115</td>
<td>H(1)</td>
<td>P225</td>
<td>none</td>
</tr>
<tr>
<td>G73</td>
<td>C(1) R(1)</td>
<td>F116</td>
<td>I(95) F(1)</td>
<td>F227</td>
<td>none</td>
</tr>
<tr>
<td>L76</td>
<td>none</td>
<td>V118</td>
<td>R(1)</td>
<td>M230</td>
<td>R(2) I(1)</td>
</tr>
<tr>
<td>V82</td>
<td>I(1)</td>
<td>Q151</td>
<td>R(1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I84</td>
<td>none</td>
<td>M184</td>
<td>I(1) L(3) T(1) L(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N88</td>
<td>none</td>
<td>L210</td>
<td>F(103)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L90</td>
<td>M(1)</td>
<td>T215</td>
<td>C(159) P(4) S(4)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NNRTI, non-nucleoside reverse transcriptase inhibitor; NRTI, nucleoside reverse transcriptase inhibitor; PI, protease inhibitor, PR, protease, RT, reverse transcriptase.
*Number of amino acid changes at designated position. Changes to amino acids conferring drug resistance shown in bold.
*Number of genomes detected by SGS with designated mutation out of 2229 genomes.

References