

**Figure 1.** Binomial Response Model Used to Evaluate Subtype and Treatment Effects on Genotypic Evolution for Each Protease and RT Position

A separate model was created for each non-B subtype. The frequencies of mutations at each position in four patient groups (untreated subtype B, treated subtype B, untreated non-B, and treated non-B) were converted to *Y* scores using a cube root transformation (similar to a logistic transform). Subtype effect was evaluated by calculating  $\theta_1$ , the score differences between non-B and B subtypes in untreated persons. The treatment effect was evaluated by calculating  $\theta_2$ , the score differences between treated and untreated persons within the same subtype. The subtype-treatment interaction was evaluated by calculating  $\theta_3$ , the difference of differences in the  $2 \times 2$  table, or the difference in treatment effects between non-B and B subtypes.  
DOI: 10.1371/journal.pmed.0020112.g001

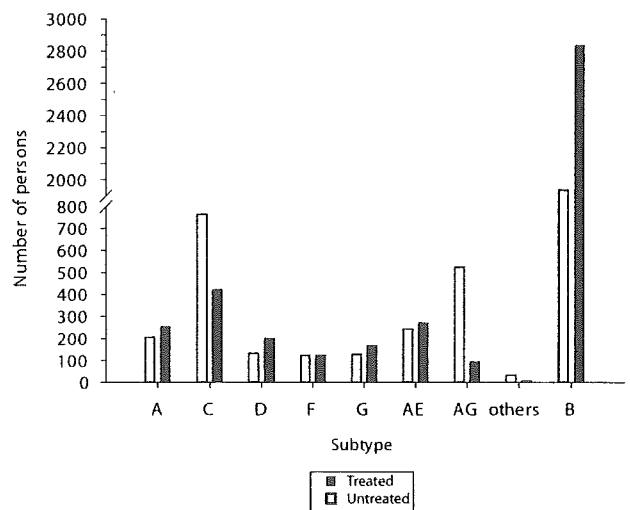
101, 103, 106, 108, 179, 181, 188, 190, 225, 227, 230, 236, and 238; and 22 protease inhibitor (PI)-resistance positions at 10, 20, 24, 30, 32, 33, 36, 46, 47, 48, 50, 53, 54, 63, 71, 73, 77, 82, 84, 88, 90, and 93 [38,39]. Mutations also included differences from consensus B that were present as part of a nucleotide mixture.

Polymorphisms were defined as mutations that occurred in more than 1% of sequences from untreated persons. Subtype-specific polymorphisms were defined as mutations that were significantly more prevalent in each non-B subtype than in subtype B viruses from untreated persons. Subtype-specific treatment-related mutations were defined as mutations that were significantly more prevalent in HIV-1 isolates from treated persons than in isolates from untreated persons infected with the same subtype.

**Quality Control**

Phylogenetic analysis to detect duplicate sequences identified 23 pairs of identical sequences and 1,039 pairs of sequences that differed from one another by no more than 1% of their nucleotides. To reduce the likelihood that isolates from different persons with similar mutations resulted from duplicate reporting, laboratory contamination, or HIV-1 transmission, only one sequence from each of these 1,062 sequence pairs was included in the analyses in this study. Although all extant HIV-1 isolates are epidemiologically linked through chains of transmission, protease or RT sequences that diverge by 1% or less appear to be more likely to result from direct transmission than those that diverge by more than 1% [40].

To distinguish mutations developing in multiple individuals from mutations that developed in a smaller number of founder viruses, we reconstructed the ancestral sequences at



**Figure 2.** Number of Treated and Untreated Persons Infected with B and Non-B HIV-1 Subtypes from Whom Protease and/or RT Sequences Were Available for Analysis.

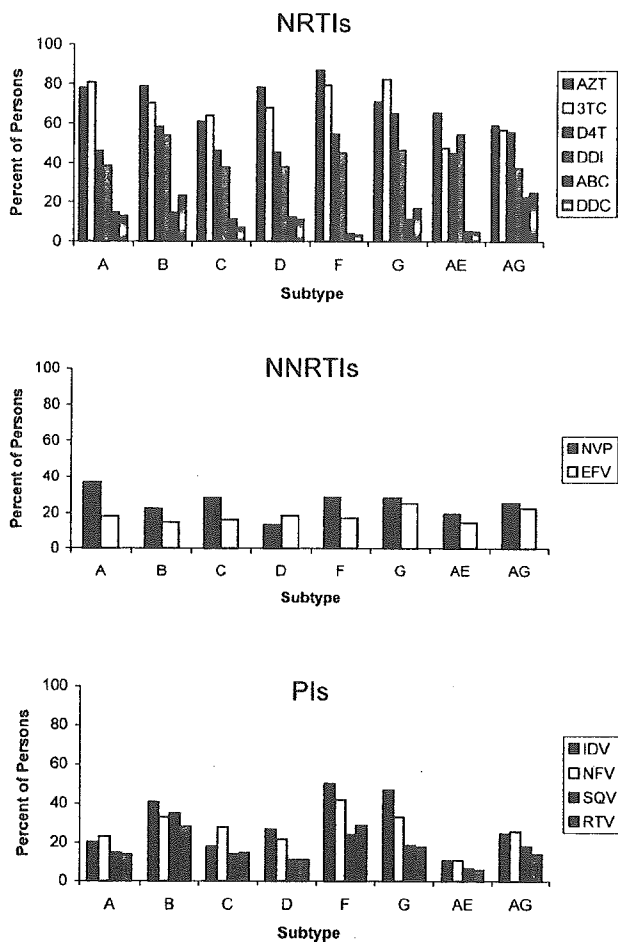
DOI: 10.1371/journal.pmed.0020112.g002

each node of the phylogenetic trees for each subtype and counted the number of times each mutation was predicted to have developed within a subtype. Mutations for which founder viruses accounted for 75% or more of occurrences (i.e., the number of branches on which the mutation has developed divided by the number of sequences with the mutation is less than 75%) were not considered treatment-

**Table 1.** Geographical Origin of Persons Infected with Non-B Subtypes

Subtype	Country
A	Uganda (215), United Kingdom (49), Kenya (41), Rwanda (34), Canada (23), France (21), Belgium (14), Sweden (10), Cameroon (10), Other (43)
C	India (225), South Africa (220), United Kingdom (145), Brazil (106), Canada (93), Zimbabwe (71), Israel (59), Botswana (42), Zambia (29), Sweden (29), Belgium (24), Uganda (19), Denmark (17), Burundi (16), France (14), Other (73)
D	Uganda (189), Cameroon (21), Cuba (21), Kenya (20), Canada (18), United Kingdom (14), Other (46)
F	Brazil (127), Romania (46), Cameroon (21), Cuba (10), Other (40)
G	Portugal (193), Cameroon (20), Spain (14), France (14), Other (53)
CRF01_AE	Thailand (260), Vietnam (98), Japan (35), Cameroon (27), Canada (11), United Kingdom (10), Uganda (10), France (10), Other (52)
CRF02_AG	Cameroon (207), Ivory Coast (128), France (47), Spain (46), Portugal (37), Senegal (24), United Kingdom (18), Italy (12), Belgium (12), Canada (11), Gabon (10), Other (63)
Others <sup>a</sup>	Cameroon (14), Other (26)

Numbers of persons from countries for which ten or more sequences were available are shown in parentheses.  
<sup>a</sup>Subtypes H, J, and K.  
DOI: 10.1371/journal.pmed.0020112.t001



**Figure 3.** Proportions of Persons Receiving Treatment with Specific NRTIs, NNRTIs, and PIs

The number of persons with non-B virus receiving the NRTI tenofovir (ten), the NNRTI delavirdine (three), and the PIs amprenavir (13) and lopinavir (28) are not shown. 3TC, lamivudine; ABC, abacavir; AZT, zidovudine; D4T, stavudine; DDC, zalcitabine; DDI, didanosine; EFV, efavirenz; IDV, indinavir; NFV, nelfinavir; NVP, nevirapine; RTV, ritonavir; SQV, saquinavir.  
DOI: 10.1371/journal.pmed.0020112.g003

related mutations. For this analysis, phylogenetic trees were created using the neighbor-joining method using the HKY85 model with gamma distribution within PAUP\* version 4.0b10 for each subtype and each gene. Ancestral sequences were reconstructed using MESQUITE version 1.02 (<http://www.mesquiteproject.org>).

To further reduce the influence of transmitted drug resistance on this analysis and to exclude the possibility that some untreated persons were classified incorrectly, sequences from untreated persons containing two or more non-

polymorphic resistance mutations were excluded from the analysis. This approach was predicated on the strong likelihood that the presence of two mutations at highly conserved sites in untreated persons does not reflect natural variation in protease or RT but rather is most consistent with previous selective drug pressure. Based on this criterion, 23 protease and 25 RT sequences from 47 persons were excluded from the analysis: 22 CRF01\_AE sequences, 15 subtype C sequences, six CRF02\_AG sequences, five subtype G sequences, four subtype A sequences, three subtype F sequences, and one subtype D sequence. The mutations in the excluded sequences consisted almost entirely of the NRTI-resistance mutations at positions 41, 67, 70, 210, 215, and 219; the NNRTI-resistance mutations at positions 103, 181, and 190; and the PI-resistance mutations at positions 48, 82, and 90. An analysis that included these 56 sequences did not affect any of the significant findings in the study because these mutations were so much more common in treated than in untreated persons in multiple different subtypes (data not shown).

### Statistical Analysis

Frequencies of mutations at each RT and protease codon were analyzed by a binomial response model employing a cube root transformation (similar to a logistic transform, with higher accuracy for extreme values) to identify significant differences in polymorphisms and treatment-related mutations between subtypes. Mutation frequencies for treated and untreated persons were compared for each subtype (Figure 1). This analysis defined three parameters: (i) a subtype parameter ( $\theta_1$ ), comparing codons between untreated persons infected with B and each of the non-B subtypes; (ii) a treatment parameter ( $\theta_2$ ), comparing codons between treated and untreated sequences of the same subtype; and (iii) an interaction parameter ( $\theta_3$ ), comparing the effect of treatment between subtype B and each of the non-B subtypes.

To increase the statistical power of our analysis, we made two simplifications. First, we did not distinguish between distinct substitutions at the same position; all differences from consensus B were considered mutations. Second, viruses were categorized only according to the classes of drugs (PI, NRTI, NNRTI) to which they had been exposed.

To correct for multiple comparisons between subtype B and each non-B subtype, significant results included those  $z$  values exceeding three in absolute value, according to a 0.05 Benjamini-Hochberg false discovery rate criterion [41]. This method is a sequential Bonferroni-type procedure that is appropriate for situations in which multiple statistically significant associations are expected. The coefficients in the binomial response model were ranked in ascending order, and each hypothesis of rank  $r$  was compared with a significance cutoff of 0.05 (false discovery rate) multiplied by  $r/n$ , where  $n$  was 99 for the protease mutations and 240 for the RT mutations (i.e., number of comparisons).

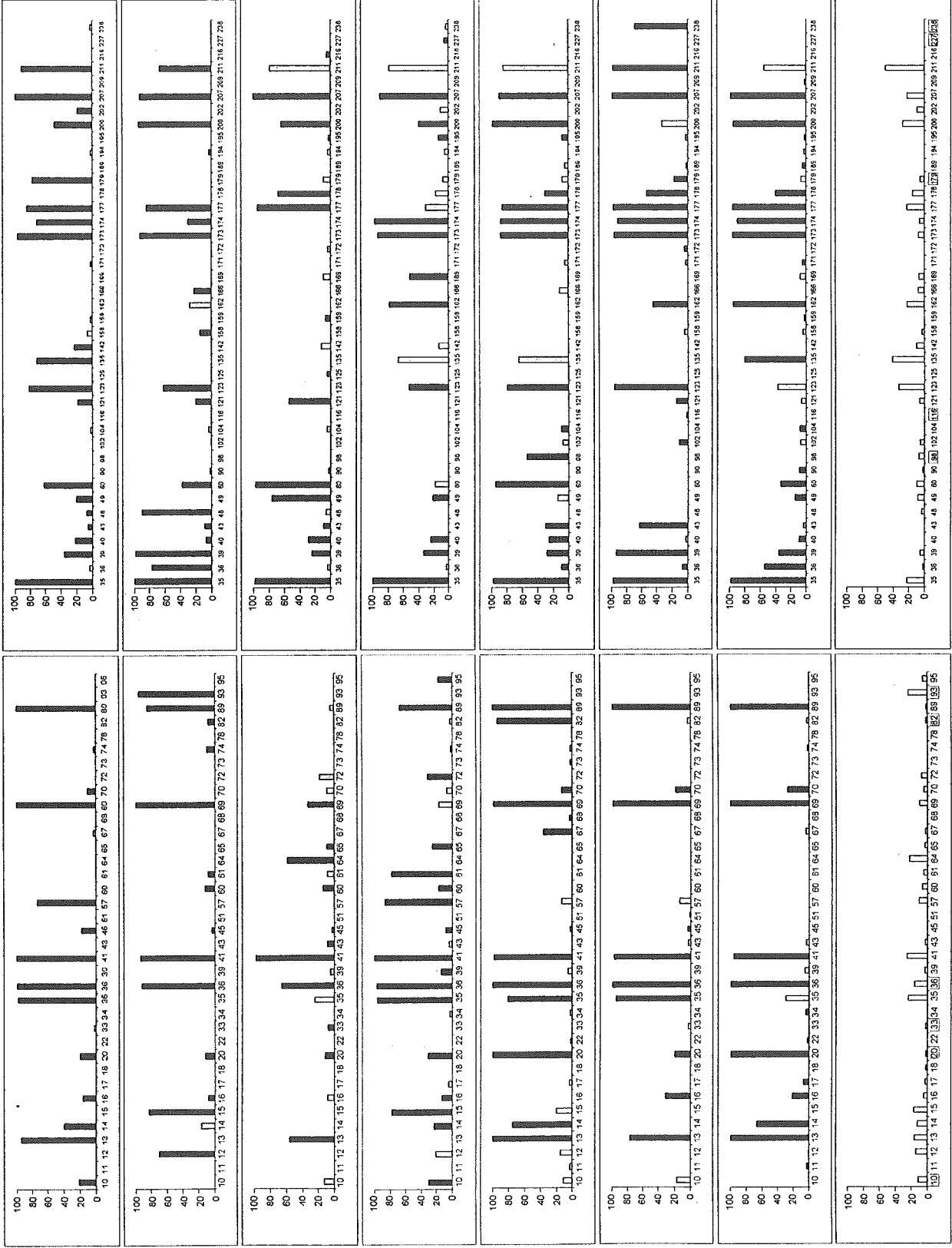
### Figure 4. Subtype-Specific Polymorphisms

Positions in protease (left) and RT (right) at which mutation frequency varied significantly between subtype B and at least one non-B subtype in untreated persons. Positions are shown along the x-axes, and the frequency of mutation for each subtype is shown along the y-axes. Positions related to drug resistance in subtype B are boxed. Bar colors denote statistical significance: black is statistically significant ( $Z_{\theta_1} \geq 3$ ); gray is borderline significant ( $1 \leq Z_{\theta_1} < 3$ ); white is not statistically significant ( $Z_{\theta_1} < 1$ ).  
DOI: 10.1371/journal.pmed.0020112.g004

Reverse Transcriptase

Polymorphisms

Protease



## Results

### HIV-1 Subtypes

Sequences were obtained from 3,686 persons, in 56 countries, infected with non-B HIV-1 subtypes (Figure 2; Table 1), including 1,997 persons from whom both protease and RT sequences were available, 908 persons from whom only protease sequences were available, and 933 persons from whom only RT sequences were available. Sequences from untreated individuals were isolated between 1983 and 2003. Sequences from treated individuals were isolated between 1993 and 2003. A total of 2,311 (82%) protease and 2,381 (83%) RT sequences were obtained from plasma samples. The remaining sequences were obtained from peripheral blood mononuclear cells.

### Antiretroviral Treatments

Of the participants with non-subtype-B viruses, 1,533 (42%) were receiving antiretroviral drugs at the time of sequencing: 1,140 were receiving NRTIs, 527 PIs, and 766 NNRTIs. According to subtype, 89% to 100% had received one or more NRTIs, 22% to 76% had received one or more PIs, and 32% to 55% had received one or more NNRTIs. Among treated persons infected with subtype B viruses, 98% had received NRTIs, 66% had received PIs, and 34% had received NNRTIs (Figure 3).

### Mutation Prevalence

Twenty-two (22%) protease and 87 (36%) RT positions were conserved in all subtypes regardless of the presence or absence of therapy. Twenty-four (24%) protease and 38 (16%) RT positions were conserved in untreated persons but were mutant in more than 1% of treated persons. The remaining 53 (53%) protease and 115 (48%) RT positions were polymorphic, or mutant in more than 1% of untreated persons.

To assess the impact of viral subtype and treatment on the distribution of mutations in protease and RT, a binomial response model using subtype and treatment as explanatory variables was used to predict whether a position was wild-type (matching the consensus B site) or mutant. This model identified three types of positions: (i) positions in sequences from untreated people more likely to be mutated in non-B than in B subtypes (subtype-specific polymorphisms); (ii) positions in sequences of the same subtype more likely to be mutated in treated than in untreated people (subtype-specific treatment-related positions); and (iii) positions for which the effect of treatment differed significantly between non-B and B subtypes (subtype-treatment interactions).

**Subtype-specific polymorphisms.** Figure 4 shows the mutation prevalence according to subtype for 37 protease and 41 RT subtype-specific polymorphisms (significant  $\theta_1$ ; see Methods). Twenty-eight of the protease and 26 of the RT subtype-

specific polymorphisms were polymorphic in untreated subtype B viruses, whereas nine of the protease and 15 of the RT were conserved in subtype B. Subtype-specific polymorphisms at conserved positions in untreated subtype B viruses were generally present in a small number of subtypes at low levels (<5%). Notable exceptions included protease positions 45 and 74 and RT positions 40, 43, 104, 195 and 238.

Six subtype-specific polymorphisms in protease (positions 10, 20, 33, 36, 82, and 93) and five in RT (98, 116, 179, 227, and 238) occurred at sites known to be associated with drug resistance in subtype B viruses. These positions (with the exception of positions 116, 227, and 238 in RT) were also polymorphic in subtype B. M184I was present in six monophyletic CRF01\_AE isolates from untreated persons [42,43] and was therefore not considered to be a subtype-specific polymorphism. These six sequences also displayed G→A hypermutation [44], possibly explaining the M184I change (ATG→ATA) and further complicating the significance of this finding.

Subtype-specific polymorphisms at four protease and three RT drug-resistance positions represented the consensus sequence for at least one non-B subtype: K20I in subtypes G and CRF02\_AG, M36I in subtypes A, C, D, F, G, CRF01\_AE and CRF02\_AG, V82I in subtype G, and I93L in subtype C for protease; and A98S in subtype G, V179I in subtype A, and K238R in CRF01\_AE for RT. Each of the non-B subtypes was significantly more polymorphic than subtype B at protease positions 20, 36, and 41 and RT positions 35, 39, and 207.

**Subtype-specific treatment-related mutations.** Figure 5 shows the mutation prevalence according to subtype for 31 protease and 36 RT subtype-specific treatment-related positions significantly more likely to be mutant in treated than untreated persons in at least one non-B subtype (significant  $\theta_2$ ; see Methods). The protease positions included 16 of the 22 known PI resistance positions and 15 additional treatment-related positions. The RT positions included 28 of the known 33 RTI resistance positions and eight additional treatment-related positions. Although each of the known PI- and RTI-resistance positions occurred in at least one non-B subtype, three of the 22 protease positions and five of the 33 RT positions included mutations that occurred too infrequently for a significant association with treatment to be detected in our analysis.

Of the 15 treatment-related protease positions not known to be associated with drug resistance, eight were also significantly associated with treatment in subtype B viruses (positions 13, 23, 43, 45, 62, 66, 74, and 85), and two have been previously reported to be associated with treatment in subtype B viruses (positions 22 and 83) [45]. The remaining five subtype-specific treatment-related protease positions included positions 6, 15, 19, 37 and 64, which—although highly polymorphic in many subtypes—are associated with treatment in subtype C (positions 6 and 64), CRF02\_AG

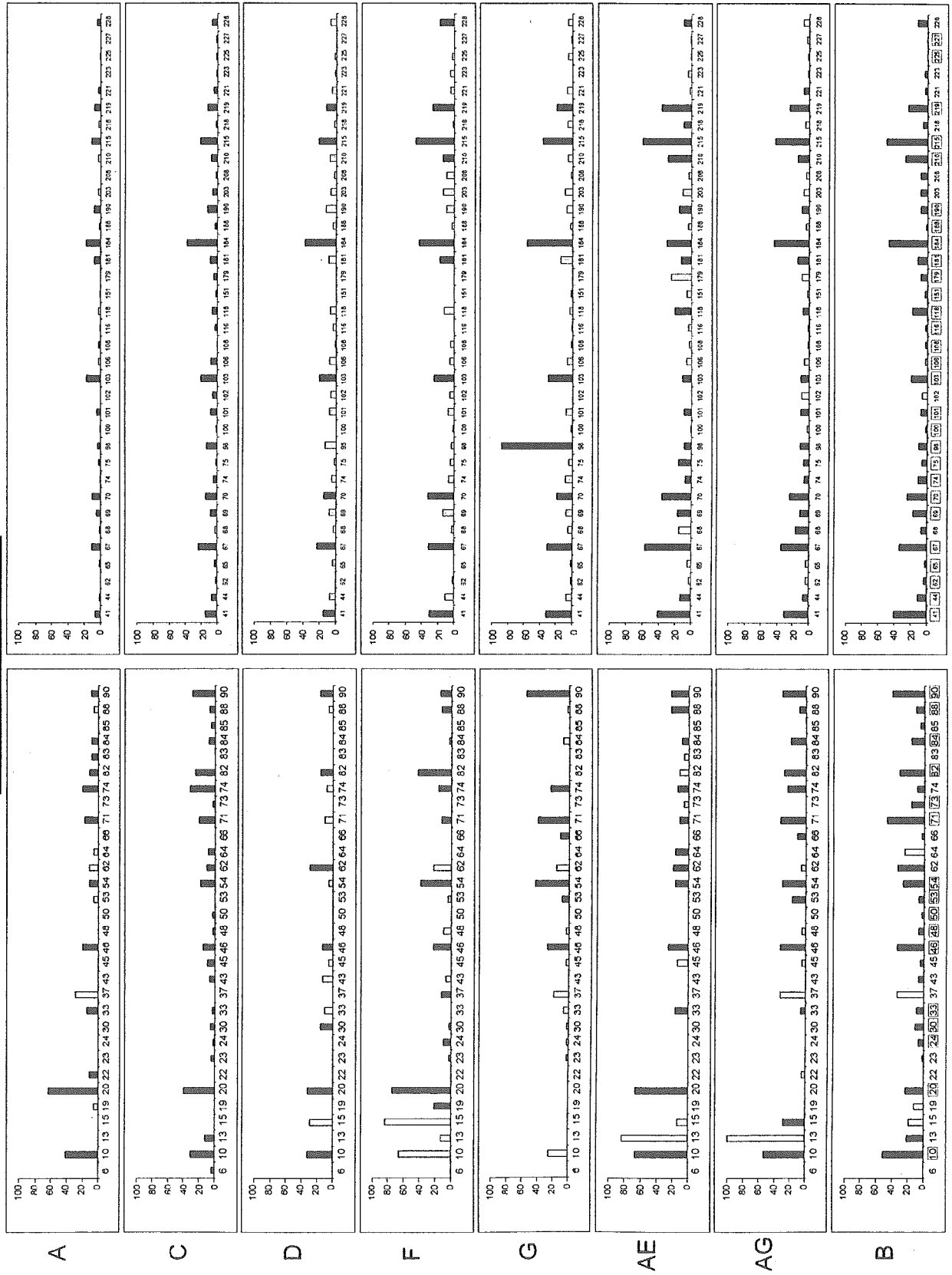
**Figure 5.** Subtype-Specific Treatment-Related Mutations

Positions in protease (left) and RT (right) at which mutations were significantly more prevalent in HIV-1 isolates from treated than from untreated persons infected with the same subtype. Positions are shown along the x-axes, and the proportion of mutant sequences in treated persons for each subtype is shown along the y-axes. For protease (left), treated persons are those receiving one or more PIs. For RT (right), treated persons are those receiving one or more NRTIs. Positions related to drug resistance in subtype B are boxed. Bar colors denote statistical significance: black is statistically significant ( $Z_{\theta_2} \geq 3$ ); gray is borderline significant ( $1 \leq Z_{\theta_2} < 3$ ); white is not statistically significant ( $Z_{\theta_2} < 1$ ). DOI: 10.1371/journal.pmed.0020112.g005

Reverse Transcriptase

Treatment Related Mutations

Protease



(position 15), subtype F (position 19), subtype A (position 37), and CRF01\_AE (position 64).

Of the eight treatment-related RT positions at sites not known to be associated with drug resistance, seven were also significantly associated with treatment in subtype B viruses (positions 68, 203, 208, 218, 221, 223, and 228) and one (position 102) was associated with treatment in subtype C but not B.

**Subtype-treatment interactions.** The subtype of the sequence significantly influenced the effect of treatment (significant  $\theta_3$ ; see Methods) on 20 protease positions (10, 12, 13, 14, 15, 20, 37, 53, 62, 63, 64, 65, 67, 71, 73, 74, 77, 82, 88, and 89) and 11 RT positions (35, 39, 48, 98, 104, 106, 121, 162, 166, 179, and 238). For example, RT position 98 was mutant in 7% of untreated and 16% of treated persons with subtype B viruses (approximately 2-fold difference) and in 1% of untreated and 14% of treated persons with CRF01\_AG viruses (14-fold difference). Other positions less likely to be mutated in subtype B than in non-B viruses in response to treatment included protease residues 14 (subtype A); 13 and 64 (subtype C); 37 and 65 (subtype F); 71 (subtype G); 62 and 64 (CRF01\_AE); and 15 and 71 (CRF02\_AG); and RT residues 35 (subtype A); 98 and 106 (subtype C); 35 and 98 (subtype G); and 98 (CRF02\_AG).

At other positions, treatment had a larger effect on subtype B viruses than on one or more non-B subtypes. For example, protease position 20 was mutant in 2% of untreated and 24% of treated persons with subtype B viruses (approximately 12-fold increase with treatment) and in 11% of untreated and 42% of treated persons with subtype C viruses (approximately 4-fold increase with treatment). These positions included protease residues 10, 20, and 63 (subtype A); 20, 53, 63, 74, and 82 (subtype C); 13 and 20 (subtype D); 10, 14, 20, and 77 (subtype F); 20, 67, 73, 82, and 88 (subtype G); 20, 63, 82, and 89 (CRF01\_AE); and 20 (CRF02\_AG); and RT residues 39 and 179 (subtype A); 35, 48, 121, and 166 (subtype C); 39 (subtype D); 39 (subtype F); 39 and 104 (subtype G); 162 and 238 (CRF01\_AE); and 39 (CRF02\_AG).

These 31 positions with subtype-treatment interactions included 12 known drug-resistance positions. Of these, seven protease (10, 20, 53, 63, 77, 82, and 88) and two RT (179 and 238) resistance positions were more likely to be mutated in subtype B than in one or more non-B subtypes in response to treatment. One protease position (71) and two RT positions (98 and 106) were more likely to be mutated in one or more non-B subtypes.

### Known Drug-Resistance Mutations

Figure 6 shows the amino acid substitutions present at drug-resistance positions in protease and RT sequences from untreated and treated persons infected with B and non-B subtypes. Fourteen of the 22 known PI-resistance positions

occurred in subtype A, 20 in subtype C, 16 in subtype D, 20 in subtype F, 18 in subtype G, 17 in CRF01\_AE, and 17 in CRF02\_AG. Thirteen of the 18 known NRTI-resistance positions occurred in subtype A, 18 in subtype C, 13 in subtype D, 15 in subtype F, 18 in subtype G, 18 in CRF01\_AE, and 16 in CRF02\_AG. Ten of the 15 known NNRTI-resistance positions occurred in subtype A, 15 in subtype C, 11 in subtype D, 13 in subtype F, 14 in subtype G, 13 in CRF01\_AE, and 12 in CRF02\_AG.

In all, 106 of 113 (94%) different amino acid substitutions at 55 known subtype B drug-resistance positions (22 protease and 33 RT) were also present in at least one non-B subtype. In an exploratory analysis, which was not controlled for multiple comparisons, the frequencies of 24 mutations at 14 protease positions and 32 mutations at 19 RT positions differed between subtype B and one or more non-B subtypes.

### Discussion

This collaborative analysis was designed to determine whether and to what degree the genetic mechanisms of HIV drug resistance are shared between subtype B and non-B viruses. Mutations responsible for drug resistance in subtype B viruses have been characterized by three types of studies: (i) those that identify mutations selected in viruses of persons receiving antiretroviral therapy, (ii) those that quantify the effect of specific mutations on *in vitro* drug susceptibility, and (iii) those that examine the effectiveness of treatment regimens in persons with viruses containing known or suspected drug-resistance mutations. This study, which identifies mutations arising in non-B viruses during antiretroviral therapy, is a necessary step for designing laboratory and clinical studies of potential drug-resistance mutations.

Do the known subtype B drug-resistance mutations also occur in non-B subtypes? We found that each of the 55 known subtype B drug-resistance mutations occurred in at least one non-B isolate. Of these, 44 (80%) were significantly associated with drug therapy in non-B isolates. The remaining 11 mutations were uncommon in subtype B and all non-B subtypes, making it difficult to determine whether they were also significantly associated with therapy. Phenotypic susceptibility testing of non-B viruses with treatment-selected mutations is necessary to confirm and quantify the contribution of these mutations to drug resistance in the genetic context in which they arise.

Do non-subtype B viruses from persons with virologic failure develop novel mutations? Fifteen protease and eight RT positions not generally considered to be drug-resistance positions were significantly associated with treatment in at least one non-B subtype. However, mutations at 17 of these 23 positions were also associated with treatment in subtype B viruses. Therefore, of the 67 mutations associated with

**Figure 6.** Amino Acid Differences from Consensus B Sequence at Drug-Resistance Positions in Protease and RT according to Subtype

(A) shows data for protease, and (B) shows data for RT. In both, the first line lists the drug-resistance positions. The second line shows single-letter amino acid codes for the consensus B sequence. For each subtype (left column), the percentage of specific mutations in untreated persons is shown above the dashed line, whereas the percentage of specific mutations in treated persons is shown below the dashed line. Positions with significant differences in mutation frequency between B and non-B subtypes ( $p < 0.01$ , according to  $\chi^2$  test with Yate's correction) are circled. A pound sign indicates an insertion.

DOI: 10.1371/journal.pmed.0020112.g006

a

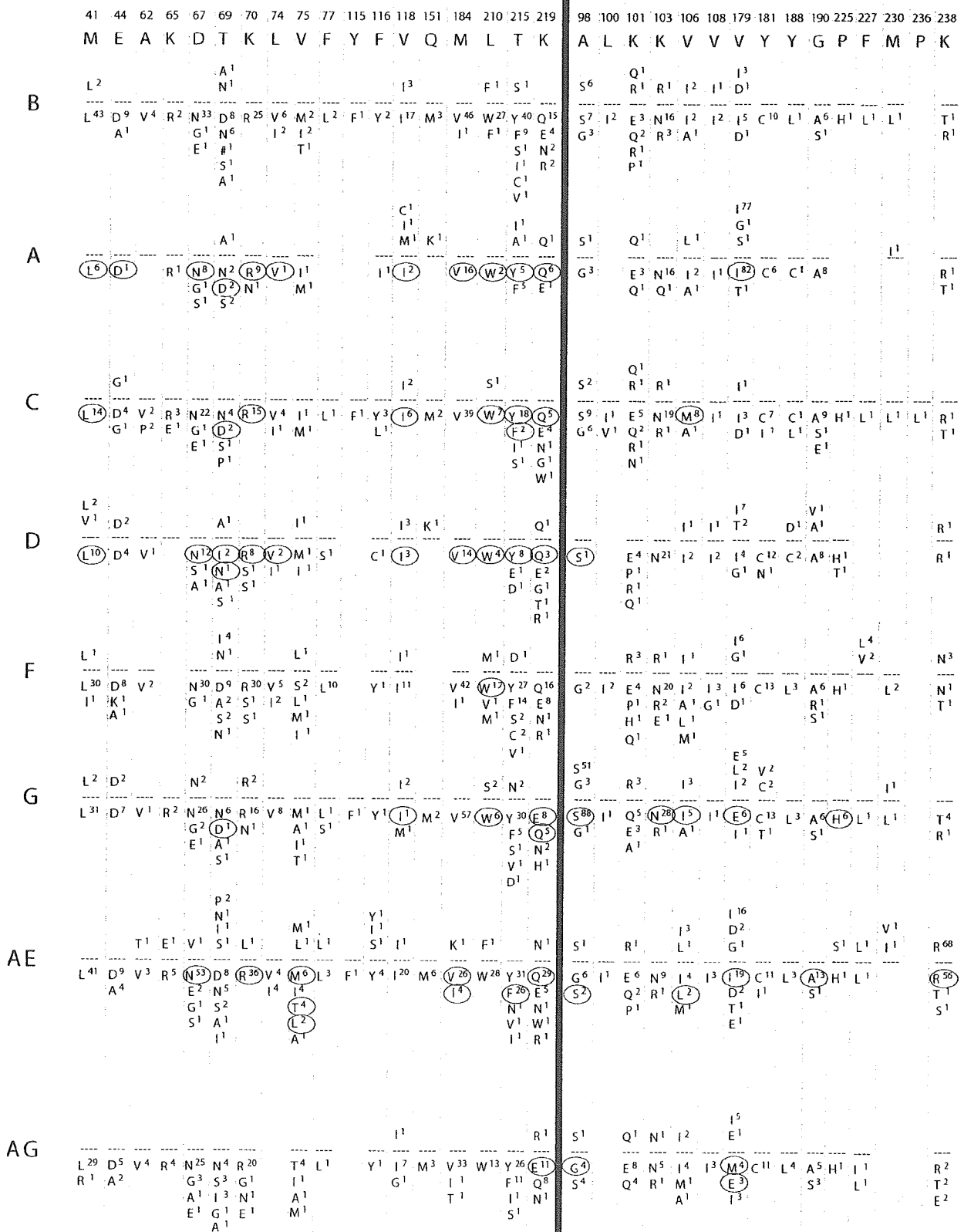
PI-related Mutations

	10	20	24	30	32	33	36	46	47	48	50	53	54	63	71	73	77	82	84	88	90	93
	L	K	L	D	V	L	M	M	I	G	I	F	I	L	A	G	V	V	I	N	L	I
<b>B</b>	I <sup>8</sup> V <sup>2</sup>	R <sup>1</sup>				V <sup>2</sup> I <sup>1</sup>	I <sup>14</sup> L <sup>1</sup>							p <sup>55</sup> S <sup>4</sup>	T <sup>5</sup> V <sup>2</sup>		I <sup>26</sup>	I <sup>2</sup>				L <sup>24</sup>
	I <sup>40</sup> V <sup>5</sup> F <sup>5</sup> R <sup>1</sup>	R <sup>10</sup> I <sup>7</sup> M <sup>3</sup> T <sup>3</sup>	I <sup>6</sup>	N <sup>10</sup>	I <sup>5</sup>	F <sup>6</sup> I <sup>2</sup> V <sup>1</sup>	I <sup>29</sup> V <sup>2</sup> L <sup>1</sup>	I <sup>22</sup> L <sup>11</sup>	V <sup>2</sup>	V <sup>6</sup>	V <sup>2</sup> L <sup>1</sup>	L <sup>6</sup> Y <sup>1</sup>	V <sup>22</sup> L <sup>2</sup> M <sup>2</sup> T <sup>2</sup>	p <sup>74</sup> A <sup>4</sup> S <sup>3</sup> T <sup>1</sup>	V <sup>33</sup> T <sup>12</sup> I <sup>3</sup>	S <sup>11</sup> C <sup>2</sup> A <sup>1</sup>	I <sup>34</sup>	A <sup>25</sup> T <sup>2</sup> I <sup>2</sup> F <sup>1</sup> S <sup>1</sup>	V <sup>16</sup>	D <sup>7</sup> S <sup>2</sup>	M <sup>41</sup>	L <sup>38</sup> M <sup>1</sup>
<b>A</b>	I <sup>13</sup> V <sup>8</sup>	R <sup>17</sup> I <sup>3</sup>				F <sup>2</sup> V <sup>1</sup>	I <sup>99</sup>	I <sup>1</sup>						p <sup>11</sup>		S <sup>1</sup>	I <sup>2</sup>	I <sup>1</sup>				L <sup>2</sup>
	I <sup>15</sup> V <sup>15</sup> F <sup>12</sup>	(I <sup>29</sup> ) R <sup>23</sup> T <sup>11</sup>		A <sup>3</sup>	I <sup>9</sup> F <sup>3</sup> V <sup>3</sup>	(I <sup>97</sup> ) L <sup>3</sup>	I <sup>14</sup> L <sup>6</sup>				L <sup>3</sup> Y <sup>3</sup>	V <sup>11</sup>	p <sup>23</sup> V <sup>3</sup> H <sup>3</sup> N <sup>3</sup> S <sup>3</sup>	V <sup>11</sup> T <sup>6</sup>				F <sup>6</sup> A <sup>3</sup> I <sup>3</sup>	V <sup>9</sup>	S <sup>6</sup>	(M <sup>9</sup> )	L <sup>9</sup>
<b>C</b>	I <sup>4</sup> V <sup>1</sup> M <sup>1</sup>	R <sup>11</sup>					I <sup>81</sup> L <sup>5</sup> V <sup>4</sup> T <sup>1</sup>							p <sup>34</sup> V <sup>6</sup> T <sup>5</sup> S <sup>5</sup>	T <sup>1</sup>		I <sup>4</sup>	I <sup>8</sup>				L <sup>95</sup>
	(I <sup>13</sup> ) F <sup>8</sup> V <sup>6</sup> P <sup>1</sup>	(R <sup>23</sup> ) T <sup>12</sup> I <sup>3</sup> M <sup>2</sup>	I <sup>2</sup>	N <sup>6</sup>	I <sup>1</sup>	F <sup>2</sup>	(I <sup>92</sup> ) L <sup>4</sup> V <sup>1</sup>	(L <sup>2</sup> )		V <sup>2</sup>	L <sup>2</sup>		V <sup>16</sup> L <sup>1</sup>	p <sup>43</sup> V <sup>9</sup> T <sup>8</sup> A <sup>4</sup>	(V <sup>19</sup> ) T <sup>3</sup> I <sup>1</sup>	(S <sup>3</sup> )	(I <sup>2</sup> )	A <sup>16</sup> I <sup>6</sup> T <sup>2</sup>	(V <sup>7</sup> ) S <sup>3</sup>	D <sup>4</sup> S <sup>3</sup>	M <sup>31</sup>	(I <sup>99</sup> ) P <sup>1</sup>
<b>D</b>	V <sup>7</sup> I <sup>5</sup>	R <sup>10</sup>				V <sup>6</sup> I <sup>1</sup>	V <sup>1</sup> L <sup>1</sup>							p <sup>32</sup> Q <sup>6</sup> S <sup>4</sup>	T <sup>2</sup>		I <sup>10</sup>	I <sup>2</sup>				L <sup>7</sup>
	(V <sup>19</sup> ) I <sup>14</sup>	R <sup>16</sup> M <sup>11</sup> T <sup>5</sup> I <sup>3</sup>		N <sup>16</sup>		V <sup>5</sup> F <sup>3</sup> I <sup>3</sup>	(I <sup>73</sup> ) L <sup>3</sup>	I <sup>8</sup> L <sup>5</sup>					V <sup>5</sup>	p <sup>51</sup> Q <sup>8</sup> T <sup>8</sup>	T <sup>8</sup> (V <sup>3</sup> )	S <sup>3</sup>	(I <sup>5</sup> )	A <sup>8</sup> I <sup>5</sup> F <sup>3</sup>	V <sup>3</sup>	D <sup>5</sup>	M <sup>16</sup>	L <sup>11</sup>
<b>F</b>	V <sup>21</sup> I <sup>9</sup> N <sup>1</sup>	R <sup>30</sup> M <sup>1</sup>					I <sup>93</sup> V <sup>2</sup> D <sup>1</sup> T <sup>1</sup> L <sup>1</sup>	I <sup>1</sup>		R <sup>1</sup>				T <sup>16</sup> P <sup>10</sup> S <sup>8</sup>	G <sup>1</sup>		I <sup>14</sup>	I <sup>2</sup>				L <sup>22</sup> V <sup>1</sup>
	(I <sup>38</sup> ) V <sup>27</sup> T <sup>1</sup>	(R <sup>64</sup> ) T <sup>9</sup> I <sup>3</sup>	I <sup>9</sup>	N <sup>3</sup>		V <sup>1</sup>	(I <sup>95</sup> ) V <sup>4</sup> L <sup>1</sup>	I <sup>13</sup> L <sup>10</sup>		V <sup>6</sup> W <sup>1</sup> R <sup>1</sup>	V <sup>1</sup>	L <sup>4</sup>	V <sup>38</sup> T <sup>3</sup>	T <sup>43</sup> P <sup>21</sup> G <sup>1</sup> I <sup>1</sup>	(V <sup>6</sup> ) T <sup>4</sup> I <sup>1</sup>	S <sup>1</sup>	(I <sup>3</sup> )	A <sup>35</sup> T <sup>4</sup> I <sup>3</sup> F <sup>1</sup>	(V <sup>3</sup> ) S <sup>10</sup> D <sup>3</sup>	(M <sup>19</sup> )	(I <sup>18</sup> )	
<b>G</b>	I <sup>9</sup> S <sup>2</sup>	I <sup>99</sup> R <sup>1</sup>	I <sup>1</sup>			I <sup>2</sup>	I <sup>100</sup>	L <sup>1</sup>	V <sup>2</sup>	W <sup>1</sup>				p <sup>13</sup> V <sup>2</sup>		R <sup>2</sup>	I <sup>3</sup>	I <sup>95</sup>		S <sup>2</sup>		L <sup>7</sup> V <sup>1</sup>
	(I <sup>24</sup> ) V <sup>1</sup>	(I <sup>94</sup> ) T <sup>5</sup>	I <sup>2</sup>	N <sup>2</sup>		I <sup>6</sup> V <sup>1</sup>	(I <sup>100</sup> )	I <sup>20</sup> L <sup>8</sup>		V <sup>2</sup> M <sup>1</sup>		L <sup>9</sup>	(V <sup>41</sup> ) A <sup>2</sup>	p <sup>34</sup> H <sup>1</sup>	V <sup>26</sup> T <sup>11</sup> I <sup>1</sup> D <sup>1</sup>		(I <sup>5</sup> )	(I <sup>64</sup> ) (T <sup>14</sup> ) (A <sup>5</sup> ) (M <sup>3</sup> ) S <sup>2</sup> F <sup>1</sup>	(V <sup>8</sup> ) S <sup>1</sup> D <sup>1</sup>	M <sup>55</sup>	(M <sup>6</sup> ) (I <sup>5</sup> )	
<b>AE</b>	I <sup>13</sup> V <sup>3</sup>	R <sup>17</sup> I <sup>3</sup>				F <sup>2</sup> V <sup>1</sup>	I <sup>100</sup>	I <sup>1</sup>						C <sup>12</sup> P <sup>12</sup>			I <sup>4</sup>	I <sup>4</sup>				L <sup>17</sup>
	(F <sup>31</sup> ) I <sup>28</sup> V <sup>8</sup>	(R <sup>31</sup> ) T <sup>19</sup> I <sup>19</sup>				(I <sup>1</sup> ) F <sup>6</sup>	(I <sup>100</sup> )	I <sup>22</sup> L <sup>3</sup>		V <sup>3</sup>		L <sup>3</sup>	V <sup>14</sup> L <sup>3</sup>	p <sup>22</sup> V <sup>6</sup>	T <sup>8</sup> (V <sup>3</sup> )	S <sup>6</sup>	(I <sup>3</sup> )	F <sup>6</sup> A <sup>3</sup> I <sup>3</sup>	V <sup>6</sup> L <sup>3</sup>	(S <sup>22</sup> )	M <sup>22</sup>	L <sup>25</sup>
<b>AG</b>		I <sup>95</sup> R <sup>2</sup> V <sup>6</sup> I <sup>5</sup> M <sup>1</sup>					I <sup>98</sup> L <sup>1</sup>							p <sup>11</sup> S <sup>3</sup>			I <sup>3</sup>	I <sup>3</sup>				F <sup>1</sup> M <sup>1</sup>
	(I <sup>28</sup> ) I <sup>21</sup> G <sup>2</sup> Y <sup>2</sup>	(I <sup>81</sup> ) R <sup>9</sup> V <sup>4</sup> T <sup>2</sup>			I <sup>4</sup>	V <sup>2</sup> F <sup>2</sup> I <sup>2</sup>	(I <sup>100</sup> )	I <sup>28</sup> L <sup>4</sup>	V <sup>4</sup>	V <sup>2</sup> A <sup>2</sup>		L <sup>17</sup>	V <sup>30</sup>	p <sup>30</sup> T <sup>4</sup> M <sup>2</sup>	V <sup>19</sup> T <sup>9</sup> I <sup>4</sup>		(I <sup>2</sup> )	A <sup>11</sup> I <sup>6</sup> F <sup>4</sup> S <sup>2</sup> T <sup>2</sup> M <sup>2</sup>	V <sup>17</sup> C <sup>2</sup>	S <sup>9</sup>	M <sup>30</sup>	

b

NRTI-related Mutations

NNRTI-related Mutations





treatment in at least one non-B subtype, 61 were also associated with treatment in subtype B. For the six mutations associated with therapy in at least one non-B subtype but not in subtype B, the associations were at the borderline of significance and require confirmation.

Among untreated persons, non-B subtype-specific polymorphisms occurred at 37 protease and 41 RT positions. Most of these non-B polymorphic positions are also polymorphic in subtype B viruses, and several act as accessory drug-resistance mutations in subtype B viruses. Phenotypic susceptibility testing of non-B viruses with such polymorphic accessory mutations is needed to confirm that these naturally occurring viruses are fully susceptible to current antiretrovirals—a supposition that appears to be true based on the excellent virologic responses of non-B viruses to antiretroviral treatment in observational studies.

We made two simplifications in this study to increase the statistical power of our analyses. These will become unnecessary in future analyses as sufficient numbers of sequences from persons with well characterized treatment histories become available. First, we did not distinguish between different substitutions at the same position; all differences from consensus B were considered mutations. Second, viruses were classified only by the classes of drugs to which they were exposed rather than by individual drugs or drug regimens. Therefore, our analyses could not detect differences between subtype B and other subtypes that depend on specific mutations or specific drugs. Indeed, two such differences have been reported: (i) V106M is the most common substitution at RT position 106 in subtype C viruses whereas V106A predominates in subtype B viruses [33,35,46], and (ii) although the protease mutations D30N and L90M both develop in non-B viruses during nelfinavir therapy, D30N occurs more commonly in subtype B, whereas L90M occurs more commonly in subtypes C, G, and CRF01\_AE [17,34,47,48].

Although the clinical samples in this study were originally obtained for a variety of purposes, including clinical management, the sequences of these samples represent experiments of nature that reveal the mutations associated with continued HIV-1 replication in the presence of selective antiretroviral therapy. The accurate identification of treatment-related mutations in such a cross-sectional analysis is challenging, however, because misclassification can result from the transmission of drug-resistant viruses, differences in specific HIV-1 variants among different human populations (population stratification), and the many statistical comparisons required as a result of HIV-1 sequence variability.

The transmission of drug-resistant HIV-1 viruses weakens cross-sectional analyses because some untreated persons may have been infected with viruses already containing treatment-related mutations. To mitigate this effect, we excluded isolates from untreated persons containing two or more non-polymorphic known drug-resistance mutations, because this pattern is not consistent with natural sequence variation. However, as noted in the Methods, an analysis that included these isolates did not alter any of the significant findings in the study. Conversely, the possibility that resistance mutations transmitted between persons in our dataset inflated the amount of resistance among persons receiving treatment was mitigated by excluding any isolate differing from another isolate at less than 1% of its nucleotides.

HIV-1 evolution is driven by genetic drift, immunologic pressure, and selective drug pressure. Population stratification can be a confounding factor when viral lineages with different founder mutations (resulting from drift or immunologic pressure) are exposed to different degrees of antiretroviral selection pressure. To distinguish mutations developing in multiple individuals as a result of selective drug pressure from mutations originating in a fewer number of founder viruses, we reconstructed the ancestral sequences at each node of a phylogenetic tree for each subtype and counted the number of times each mutation was predicted to have developed within that subtype. Because of the limited ability of phylogenetic methods to estimate accurate trees for large numbers of related sequences (i.e., belonging to the same HIV-1 subtype), only those positions for which the majority of mutations ( $\geq 75\%$ ) appeared to result from new mutations were considered to be selected by antiretroviral therapy.

Because this analysis was, to our knowledge, the first to simultaneously examine all protease and most polymerase-coding RT positions in multiple subtypes, and because multiple associations between mutation and treatment were expected, we used a relatively lenient correction for multiple comparisons in order to minimize the number of missed associations. Nonetheless, of the 67 positive associations detected in this study, 61 were also present in persons with subtype B viruses and have previously been reported [45,49].

In conclusion, most of the protease and RT positions associated with drug resistance in subtype B viruses are selected by antiretroviral therapy in one or more non-B subtypes as well. Conversely, we found no evidence that non-B viruses develop resistance by mutations at positions that are not associated with resistance in subtype B viruses. Based on currently available data, global surveillance efforts and genotypic assessments of drug resistance should focus primarily on the known subtype B drug-resistance mutations.

## Supporting Information

**Dataset S1.** List of GenBank Accession Numbers for Non-Subtype-B Sequences Used in This Study

Found at DOI: 10.1371/journal.pmed.0020112.s001 (59 KB PDF).

### Accession Numbers

The GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/>) isolates and accession numbers for the reference subtype specimens discussed in this paper are U455/subtype A (M62320), HXB2/subtype B (K03455), C2220/subtype C (U46016), NDK/subtype D (M27323), 93BR020/subtype E (AF005494), SE6165/subtype G (AF061642), 90CR056/subtype H (AF005496), SE9173c/subtype J (AF082394), 97EQT11C/subtype K (AJ249235), CM240/CRF01\_AE (U54771), 1bNG/CRF02\_AG (L39106), YBF30/Group N (AJ006022), and AN170C/Group O (L20587). The accession numbers for the non-subtype-B sequences used in this study are listed in Dataset S1.

## Acknowledgments

D. Katzenstein is the recipient of a Doris Duke Distinguished Clinical Scientist award, and the support of the Doris Duke Charitable Foundation for R. Kantor and the project is gratefully acknowledged. R. Shafer was supported by grants from the National Institutes of Allergy and Infectious Diseases (AI46148-01) and National Institute of General Medical Sciences (5P01GM066524-02).

We thank Paul Weidle for data submission. We thank Matthew J. Gonzales, Soo-Yon Rhee, and Tommy Liu for computation assistance and Jonathan Taylor for help with the statistical analysis. The Doris Duke Charitable Foundation did not have any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Author contributions.** R. Kantor, D. A. Katzenstein, B. Efron, J. M. Schapiro, and R. W. Shafer designed the study. R. Kantor, D. A. Katzenstein, B. Efron, A. P. Carvalho, B. Wynhoven, P. Cane, J. Clarke, S. Sirivichayakul, M. A. Soares, J. Snoeck, C. Pillay, H. Rudich, R. Rodrigues, A. Holguin, K. Ariyoshi, M. B. Bouzas, and R. W. Shafer analyzed the data. R. Kantor, D. A. Katzenstein, P. Cahn, W. Sugiura, V. Soriano, L. F. Brigido, Z. Grossman, L. Morris, A.-M. Vandamme, A. Tanuri, P. Phanuphak, J. N. Weber, D. Pillay, P. R. Harrigan, R. Camacho, and R. W. Shafer interpreted the data. R. Kantor, D. A. Katzenstein, and R. W. Shafer contributed to writing the paper. ■

## References

- Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, et al. (1999) Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* 397: 436–441.
- Hahn BH, Shaw GM, De Cock KM, Sharp PM (2000) AIDS as a zoonosis: Scientific and public health implications. *Science* 287: 607–614.
- Korber B, Muldoon M, Theiler J, Gao F, Gupta R, et al. (2000) Timing the ancestor of the HIV-1 pandemic strains. *Science* 288: 1789–1796.
- Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, et al. (2000) HIV-1 nomenclature proposal. *Science* 288: 55–56.
- Osmanov S, Patou G, Walker N, Schwardlander B, Esparza J (2002) Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000. *J Acquir Immune Defic Syndr* 29: 184–190.
- Gonzales MJ, Machekeano RN, Shafer RW (2001) HIV-1 reverse transcriptase and protease subtypes: Classification, amino acid mutation patterns, and prevalence in a Northern California clinic-based population. *J Infect Dis* 184: 998–1006.
- Dumans AT, Soares MA, Machado ES, Hue S, Brindeiro RM, et al. (2004) Synonymous genetic polymorphisms within Brazilian human immunodeficiency virus type 1 subtypes may influence mutational routes to drug resistance. *J Infect Dis* 189: 1232–1238.
- Gonzalez LM, Brindeiro RM, Aguiar RS, Pereira HS, Abreu CM, et al. (2004) Impact of nelfinavir resistance mutations on in vitro phenotypic fitness, and replication capacity of human immunodeficiency virus type 1 with subtype B and C proteases. *Antimicrob Agents Chemother* 48: 3552–3555.
- Shafer RW, Eisen JA, Merigan TC, Katzenstein DA (1997) Sequence and drug susceptibility of subtype C reverse transcriptase from human immunodeficiency virus type 1 seroconverters in Zimbabwe. *J Virol* 71: 5441–5448.
- Shafer RW, Chuang TK, Hsu P, White CB, Katzenstein DA (1999) Sequence and drug susceptibility of subtype C protease from human immunodeficiency virus type 1 seroconverters in Zimbabwe. *AIDS Res Hum Retroviruses* 15: 65–69.
- Toni T, Masquelier B, Bonard D, Faure M, Huer C, et al. (2002) Primary HIV-1 drug resistance in Abidjan (Cote d'Ivoire): A genotypic and phenotypic study. *AIDS* 16: 488–491.
- Palmer S, Margot N, Gilbert H, Shaw N, Buckheit R Jr, et al. (2001) Tenofovir, adefovir, and zidovudine susceptibilities of primary human immunodeficiency virus type 1 isolates with non-B subtypes or nucleoside resistance. *AIDS Res Hum Retroviruses* 17: 1167–1173.
- Aperrei C, Descamps D, Gollin G, Loussert-Ajaka I, Damond F, et al. (1998) Human immunodeficiency virus type 1 subtype F reverse transcriptase sequence and drug susceptibility. *J Virol* 72: 3534–3538.
- Adje C, Cheingsong R, Roels TH, Maurice G, Djomand G, et al. (2001) High prevalence of genotypic and phenotypic HIV-1 drug-resistant strains among patients receiving antiretroviral therapy in Abidjan, Cote d'Ivoire. *J Acquir Immune Defic Syndr* 26: 501–506.
- Weidle PJ, Malamba S, Mwebaze R, Sozi C, Rukundo G, et al. (2002) Assessment of a pilot antiretroviral drug therapy programme in Uganda: Patients' response, survival, and drug resistance. *Lancet* 360: 34–40.
- Vergne L, Malonga-Mouellet G, Mistroul I, Mavoungou R, Mansaray H, et al. (2002) Resistance to antiretroviral treatment in Gabon: Need for implementation of guidelines on antiretroviral therapy use and HIV-1 drug resistance monitoring in developing countries. *J Acquir Immune Defic Syndr* 29: 165–168.
- Cane PA, de Ruiter A, Rice PA, Wiselka M, Fox R, et al. (2001) Resistance-associated mutations in the human immunodeficiency virus type 1 subtype C protease gene from treated and untreated patients in the United Kingdom. *J Clin Microbiol* 39: 2652–2654.
- Pillay D, Walker AS, Gibb DM, De Rossi A, Kaye S, et al. (2002) Impact of human immunodeficiency virus type 1 subtypes on virologic response and emergence of drug resistance among children in the Paediatric European Network for Treatment of AIDS (PENTA) 5 trial. *J Infect Dis* 186: 617–625.
- Frater AJ, Dunn DT, Beardall A, Ariyoshi K, Clarke JR, et al. (2002) Comparative response of African HIV-1-infected individuals to highly active antiretroviral therapy. *AIDS* 16: 1139–1146.
- Grossman Z, Vardinon N, Chemtob D, Alkan ML, Bentwich Z, et al. (2001) Genotypic variation of HIV-1 reverse transcriptase and protease: Comparative analysis of clade C and clade B. *AIDS* 15: 1453–1460.
- Kebba A, Arwine D, Mwebaze R, Kityo C, Nakityo R, et al. (2002) Therapeutic responses to AZT + 3TC + EFV in advanced antiretroviral naive HIV type 1-infected Ugandan patients. *AIDS Res Hum Retroviruses* 18: 1181–1187.
- Landman R, Schiemann R, Thiam S, Vray M, Canestri A, et al. (2003) Once-a-day highly active antiretroviral therapy in treatment-naive HIV-1-infected adults in Senegal. *AIDS* 17: 1017–1022.
- Alexander CS, Montessori V, Wynhoven B, Dong W, Chan K, et al. (2002) Prevalence and response to antiretroviral therapy of non-B subtypes of HIV in antiretroviral-naive individuals in British Columbia. *Antivir Ther* 7: 31–35.
- Weidle PJ, Downing R, Sozi C, Mwebaze R, Rukundo G, et al. (2003) Development of phenotypic and genotypic resistance to antiretroviral therapy in the UNAIDS HIV Drug Access Initiative—Uganda. *AIDS* 17 (Suppl 3): S39–S48.
- Djomand G, Roels T, Ellerbrock T, Hanson D, Diomande F, et al. (2003) Virologic and immunologic outcomes and programmatic challenges of an antiretroviral treatment pilot project in Abidjan, Cote d'Ivoire. *AIDS* 17 (Suppl 3): S5–S15.
- Laurent C, Kouanfack C, Koulla-Shiro S, Nkoue N, Bourgeois A, et al. (2004) Effectiveness and safety of a generic fixed-dose combination of nevirapine, stavudine, and lamivudine in HIV-1-infected adults in Cameroon: Open-label multicentre trial. *Lancet* 364: 29–34.
- Palmer S, Alaeus A, Albert J, Cox S (1998) Drug susceptibility of subtypes A, B, C, D, and E human immunodeficiency virus type 1 primary isolates. *AIDS Res Hum Retroviruses* 14: 157–162.
- Loemba H, Brenner B, Parniak MA, Ma'ayan S, Spira B, et al. (2002) Genetic divergence of human immunodeficiency virus type 1 Ethiopian clade C reverse transcriptase (RT) and rapid development of resistance against nonnucleoside inhibitors of RT. *Antimicrob Agents Chemother* 46: 2087–2094.
- Caride E, Brindeiro R, Hertogs K, Larder B, Dehertogh P, et al. (2000) Drug-resistant reverse transcriptase genotyping and phenotyping of B and non-B subtypes (F and A) of human immunodeficiency virus type 1 found in Brazilian patients failing HAART. *Virology* 275: 107–115.
- Caride E, Hertogs K, Larder B, Dehertogh P, Brindeiro R, et al. (2001) Genotypic and phenotypic evidence of different drug-resistance mutation patterns between B and non-B subtype isolates of human immunodeficiency virus type 1 found in Brazilian patients failing HAART. *Virus Genes* 23: 193–202.
- Velazquez-Campoy A, Vega S, Freire E (2002) Amplification of the effects of drug resistance mutations by background polymorphisms in HIV-1 protease from African subtypes. *Biochemistry* 41: 8613–8619.
- Ariyoshi K, Matsuda M, Miura H, Tateishi S, Yamada K, et al. (2003) Patterns of point mutations associated with antiretroviral drug treatment failure in CRF01\_AE (subtype E) infection differ from subtype B infection. *J Acquir Immune Defic Syndr* 33: 336–342.
- Brenner B, Turner D, Oliveira M, Moisi D, Deterio M, et al. (2003) A V106M mutation in HIV-1 clade C viruses exposed to efavirenz confers cross-resistance to non-nucleoside reverse transcriptase inhibitors. *AIDS* 17: F1–F5.
- Grossman Z, Paxinos EE, Averbuch D, Maayan S, Parkin NT, et al. (2004) Mutation D30N is not preferentially selected by human immunodeficiency virus type 1 subtype C in the development of resistance to nelfinavir. *Antimicrob Agents Chemother* 48: 2159–2165.
- Grossman Z, Istomin V, Averbuch D, Lorber M, Risenberg K, et al. (2004) Genetic variation at NNRTI resistance-associated positions in patients infected with HIV-1 subtype C. *AIDS* 18: 909–915.
- Lole KS, Bollinger RC, Paranjape RS, Gadhari D, Kulkarni SS, et al. (1999) Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 73: 152–160.
- Shafer RW, Jung DR, Bets BJ (2000) Human immunodeficiency virus type 1 reverse transcriptase and protease mutation search engine for queries. *Nat Med* 6: 1290–1292.
- D'Aquila RT, Schapiro JM, Brun-Vezinet F, Clotet B, Conway B, et al. (2003) Drug resistance mutations in HIV-1. *Top HIV Med* 11: 92–96.
- Rhee SY, Gonzales MJ, Kantor R, Bets BJ, Ravela J, et al. (2003) Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res* 31: 298–303.
- Hue S, Clewley JP, Cane PA, Pillay D (2004) HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. *AIDS* 18: 719–728.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 57: 289–300.
- Lan NT, Recordon-Pinson P, Hung PV, Uyen NT, Lien TT, et al. (2003) HIV type 1 isolates from 200 untreated individuals in Ho Chi Minh City (Vietnam): ANRS 1257 study. Large predominance of CRF01\_AE and presence of major resistance mutations to antiretroviral drugs. *AIDS Res Hum Retroviruses* 19: 925–928.
- Tovanabutra S, Beyrer C, Sakkhachornphop S, Razak MH, Ramos GL, et al.

- (2004) The changing molecular epidemiology of HIV type 1 among northern Thai drug users, 1999 to 2002. *AIDS Res Hum Retroviruses* 20: 465–475.
44. Rose PP, Korber BT (2000) Detecting hypermutations in viral sequences with an emphasis on G → A hypermutation. *Bioinformatics* 16: 400–401.
  45. Wu TD, Schiffer CA, Gonzales MJ, Taylor J, Kantor R, et al. (2003) Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments. *J Virol* 77: 4836–4847.
  46. Morris L, Pillay C, Chezzi C, Lupondwana P, Ntsala M, et al. (2003) Low frequency of the V106M mutation among HIV-1 subtype C-infected pregnant women exposed to nevirapine. *AIDS* 17: 1698–1700.
  47. Gomes P, Diogo I, Goncalves MF, Carvalho P, Cabanas J, et al. (2002) Different pathways to nelfinavir genotypic resistance in HIV-1 subtypes B and G [abstract]. Ninth Conference on Retroviruses and Opportunistic Infections; 2002 February 24–28; Seattle, Washington. Abstract number 46. Available: <http://www.retroconference.org/2002/Abstract/12937.htm>. Accessed 14 March 2005.
  48. Sugiura W, Matsuda Z, Yokomaku Y, Hertogs K, Larder B, et al. (2002) Interference between D30N and L90M in selection and development of protease inhibitor-resistant human immunodeficiency virus type 1. *Antimicrob Agents Chemother* 46: 708–715.
  49. Gonzales MJ, Wu TD, Taylor J, Belitskaya I, Kantor R, et al. (2003) Extended spectrum of HIV-1 reverse transcriptase mutations in patients receiving multiple nucleoside analog inhibitors. *AIDS* 17: 791–799.

### Patient Summary

**Background** There are many different subtypes of HIV-1. The most common one in more developed countries is subtype B and that is the one which has been studied most and used in drug development. However, worldwide other subtypes are more frequent. All HIV subtypes acquire mutations and some of these cause resistance to the drugs used to treat HIV. It is not clear whether the same mutations that cause drug resistance to subtype B virus are also important in causing resistance to non-subtype B viruses.

**What Did the Researchers Do?** They compared the viral sequences of 3,686 people with non-subtype B HIV, and 4,769 with subtype B virus, all with known treatment histories. They found that the mutations known to cause drug resistance in subtype B virus also occur in non-subtype B, and the majority of mutations in non-subtype B also occur in subtype B.

**What Do These Findings Mean?** It seems that largely the same mutations occur in both subtype B and non-subtype B viruses. However, some mutations were only present in low numbers, so more work will need to be done before their role is clear. Also, the authors did not look at all mutations and their relation to each different drug a patient had, only the general type of drug. Nor did they look at what happens when different mutations occur at one place in a virus. However, for now, the current strategy of focusing on assessing the mutations seen in subtype B virus seems a reasonable approach to take when assessing surveillance of drug resistance, while more work is done to follow up these findings.

**Where Can I Get More Information?** TheBody.com has a section on drug resistance: <http://www.thebody.com/treat/resistance.htm>. The Aidsmap Web site has many patient information sheets, including on resistance: <http://www.aidsmap.com>.

## Role of the Specific Amino Acid Sequence of the Membrane-Spanning Domain of Human Immunodeficiency Virus Type 1 in Membrane Fusion

Kosuke Miyauchi,<sup>1</sup> Jun Komano,<sup>1</sup> Yoshiyuki Yokomaku,<sup>2</sup> Wataru Sugiura,<sup>3</sup> Naoki Yamamoto,<sup>1,3</sup> and Zene Matsuda<sup>1\*</sup>

Laboratory of Virology and Pathogenesis<sup>1</sup> and Therapeutic Research and Clinical Science Group,<sup>3</sup> AIDS Research Center, National Institute of Infectious Diseases, Tokyo, and Division of Control and Treatment of Infectious Diseases, Chiba University Hospital, Chiba,<sup>2</sup> Japan

Received 16 July 2004/Accepted 6 December 2004

**Fusion between cell and virus membranes mediated by gp41 initiates the life cycle of human immunodeficiency virus type 1. In contrast to the many studies that have elucidated the structure-function relationship of the ectodomain, the study of the membrane-spanning domain (MSD) has been rather limited. In particular, the role that the MSD's specific amino acid sequences may have in membrane fusion as well as other gp41 functions is not well understood. The MSD of gp41 contains well-conserved glycine residues that form the GXXXG motif (G, glycine; X, other amino acid residues), a motif often found at the helix-helix interface of membrane spanning  $\alpha$ -helices. Here we examined the role that the specific amino acid sequence of the gp41 MSD has in gp41 function, particularly in membrane fusion, by making two types of MSD mutants: (i) glycine substitution mutants in which glycine residues of the MSD were mutated to alanine or leucine residues, and (ii) replacement mutants in which the entire MSD was replaced with one derived from glycoporphin A or from vesicular stomatitis virus G. The substitution of glycines did not affect gp41 function. MSD-replacement mutants, however, showed severely impaired fusion activity. The assay using the Env expression vector revealed defects in membrane fusion after CD4 binding steps in the MSD-replacement mutants. In addition, the change in Env processing was noted for MSD-replacement mutants. These results suggest that the MSD of gp41 has a relatively wide but not unlimited tolerance for mutations and plays a critical role in membrane fusion as well as in other steps of Env biogenesis.**

The envelope glycoprotein (Env) of human immunodeficiency virus type 1 (HIV-1) plays a critical role in the entry process in the viral life cycle. Env is synthesized as a precursor, gp160, and then processed into a heterodimer consisting of gp120 and gp41. Interaction of gp120 with CD4 and chemokine receptors triggers the membrane fusion process.

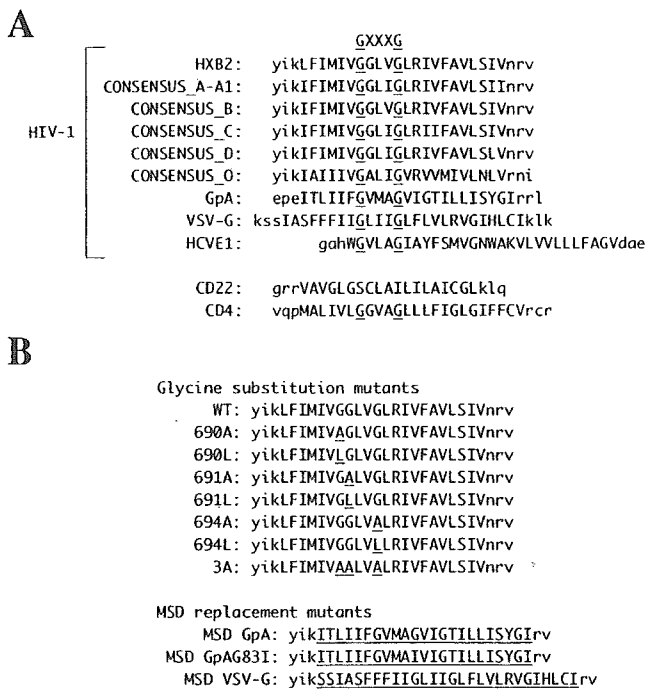
The gp41 is believed to play a central role in the fusion process during postreceptor binding. It is divided into three subunits or domains: the extracellular, membrane-spanning, and cytoplasmic domains. The contribution of the extracellular domain to membrane fusion has been well documented. It contains conserved heptad repeats preceded by the fusion peptide and is thought to undergo conformational changes during membrane fusion to form a trimeric coiled-coil, commonly observed in the envelope proteins of viruses such as the influenza virus, the Moloney murine leukemia virus, and the Ebola virus (4–6, 11, 20, 37, 43, 44).

The cytoplasmic domain is relatively long compared with those of other simple animal retroviral envelope proteins and bears two well-conserved amphipathic helices called LLP1 and LLP2 (10, 25, 40). The cytoplasmic domain is important to intracellular trafficking as well as to the efficient incorporation

of Env onto the budding viral particle (9, 32, 46). The cytoplasmic domain may affect fusion activity of Env (13, 27).

The membrane-spanning domain (MSD) of gp41 anchors Env on the lipid bilayer, and its amino acid sequences are highly conserved among independent isolates of HIV-1. Several studies have indicated that the MSD is involved in membrane fusion—the glycosylphosphatidylinositol-anchored Env of HIV-1, which lacks the MSD and the cytoplasmic domain, could not induce syncytia (35, 42). However, whether specific amino acid sequences in the MSD are required for its function in HIV-1 replication remains controversial. For example, the previous study showed that substituting a leucine residue for the conserved arginine residue within the MSD resulted in a replication-incompetent virus (31). Substituting isoleucine for the same arginine residue, however, did not affect gp41 function (42). Furthermore, fusion activity was retained in a mutant in which the entire MSD and cytoplasmic domain were replaced with those of CD4 (41). Wilk et al. have reported a replication-competent recombinant HIV-1 in which a CD22 MSD replaced the gp41 MSD (45). Because there is no apparent sequence homology among the MSDs of gp41, CD4, and CD22, these results suggest that gp41 function may not require a specific sequence in the MSD or, alternatively, that some as-yet-undetermined characteristic of CD4 and CD22 MSDs might compensate for the naturally occurring sequence. One feature common to the MSDs of gp41, CD4, and CD22 is the presence of several glycine residues (Fig. 1A). A glycine residue is not a rarity in the MSDs of membrane proteins (8, 39).

\* Corresponding author. Mailing address: Laboratory of Virology and Pathogenesis, AIDS Research Center, National Institute of Infectious Diseases, 4-7-1 Gakuen Musashimurayama, Tokyo 208-0011, Japan. Phone: 81-42-561-0771-335. Fax: 81-42-562-7875. E-mail: z matsuda@nih.go.jp.



**FIG. 1.** Amino acid sequences of the MSD of HIV-1 isolates and several membrane proteins. (A) Amino acid sequences of the predicted MSD of HXB2, HIV-1 isolates, and several membrane proteins are shown. The sequence of HIV-1 isolates was according to HIV Sequence Compendium 2001 (17). The MSDs of CD4 and CD22 that could replace the MSD of gp41 without affecting gp41's function are also shown (41, 45). The capital letters indicate the amino acid sequences within the predicted MSD. The small letters indicate the amino acid sequences surrounding the MSD. Glycine residues in the GXXXG motif are underlined. (B) Amino acid sequences of MSD mutants. The mutated portions are underlined.

With respect to several viral membrane proteins, glycine residues are more often found in the MSDs of envelope proteins than in the MSDs of nonenvelope proteins (8). Although the importance of glycine residues in Env-mediated membrane fusion has been studied with viruses such as vesicular stomatitis virus (VSV) and the influenza virus (8, 23), the role of glycine residues in the MSD of gp41 has not been investigated.

In the gp41 MSD of HIV-1, the glycine residues form the GXXXG motif (G, glycine; X, other amino acid residues). This motif is often observed in transmembrane  $\alpha$ -helices and is believed to stabilize helix-helix interactions of membrane proteins (12, 34). In the case of glycophorin A (GpA) MSD, this motif is critical for homodimerization (26). A recent study showed that the transmembrane domains of hepatitis C virus E1 and E2 envelope glycoproteins are required in heterodimerization and that E1 also has the GXXXG motif in its MSD (29). The above-mentioned CD4 also contains the GXXXG motif. Thus, the maintenance of fusion activity in gp41 mutants having MSDs that were replaced with those of CD4 or CD22 might depend on the presence of glycine residues in the CD4 and CD22 MSDs (41, 45). To investigate whether the specific amino acid sequence in the MSD of gp41 is required for fusion activity, we mutated glycine residues or replaced the entire MSD of gp41 with heterologous MSDs and analyzed the effect that these mutations had on gp41 function.

We found that the latter heterologous replacement enhanced the processing of gp160. The fusion activity of gp41 was severely impaired. Our analysis of the membrane fusion process of these mutants revealed that the defects are manifested in the postreceptor binding steps preceding lipid mixing and possibly in the steps after the initial pore formation.

## MATERIALS AND METHODS

**Construction of plasmids.** The 1.2-kb NheI-BamHI fragment covering the *env* portion of the modified HXB2, HXB2RU3ΔN (*vpr*<sup>+</sup>, *vpu*<sup>+</sup>, *nef*<sup>+</sup>, and one NheI site within the vector were deleted), was subcloned into pGEM7zf(+) (Promega, Madison, Wis.) as a target for mutagenesis. To generate each glycine substitution (Gly<sup>sub</sup>) mutant, site-directed mutagenesis was performed using a QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, Calif.) with complementary oligonucleotide pairs (for 690A, CATAATGATAGTAGCAGGCTTGGT AGGT and ACCTACCAAGCCTGCTACTATCATTATG; for 690L, CATAA TGATAGTACTAGGCTTGGTAGGT and ACCTACCAAGCCTAGTACTACT CATTATG; for 691A, ATGATAGTAGGAGCCTTGGTAGGTTTA and TAA ACCTACCAAGCCTCCTACTACTAT; for 691L, ATGATAGTAGGACTCT TGGTAGGTTTA and TAAACCTACCAAGAGTCTACTACTAT; for 694A, AGGAGGCTTGGTAGCTTAAAGAATAGTTTTT and CAAAACTATT TAAAGTACTACCAAGCCTCCT; and for 694L, AGGAGGCTTGGTAGCTT TAAAGAATAGTTTTT and CAAAACTATTCTTAAAAAGTACCAAGCCT CCT), generating the substitution of an alanine or a leucine residue for a glycine residue. The PCR was performed using Pfu turbo (Stratagene). The 3A mutant was created by site-directed mutagenesis using a complementary oligonucleotide pair (CATAATGATAGTAGCCGCTTGGTAGCCITAAAGAATAGTTTTT and CAAAACTATTCTTAAAGGCTACCAAGCGGCTACTACTACTATG).

To generate the MSD-replacement mutants, megaprimer that were produced by PCR that targeted the MSD portion of VSV-G or GpA as a template were used as mutagenesis oligonucleotides (for MSD GpA, CAAATTGGCTGTGG TATATAAAAAATCACCTGATCATC and GAATATCCCTGCCTAACTCT GATGCCGTAGCTGAT; for MSD VSV-G, CAAATTGGCTGTGGTATATA AAAAGCTCTATTGCCTC and GAATATCCCTGCCTAACTCTAATGCAA AGATGGATAC). MSD GpAG83I was produced by site-directed mutagenesis using a subclone of MSD GpA as a template with a primer set (CATCTTCGG CGTGATGGCCATCGTGATCGGCACCATCCTG and CAGGATGGTGCC GATCACGATGGCCATCACGCCGAAGATG). Following mutagenesis, the 1.2-kb NheI-BamHI fragments were sequenced and cloned back into the pSP65HXB2RU3ΔN plasmid. After the cloning back, the entire NheI-BamHI portion together with the junction was verified by sequencing.

An Env expression vector, pElucEnv, was used to express *env* genes in this study. Although the details of its construction will be described elsewhere, it is a derivative of pSP65HXB2RU3ΔN and lacks the *gag* and *pol* portions of the provirus. The *rev* function was provided by *rev* cDNA cloned into the *nef* region. The original *rev* gene was inactivated by mutating its initiation codon. A BsiWI site was introduced near the initiation codon of the original *rev* gene by this mutagenesis. Because this vector contains an enhanced green fluorescent protein (EGFP)-firefly luciferase expression module outside the provirus, the transfected cells fluoresce green, allowing us to estimate transfection efficiency by measuring firefly luciferase activity. To generate each Env expression vector, the NheI-BamHI fragment of the pSP65HXB2RU3ΔN wild type (WT), MSD GpA, MSD GpAG83I, or MSD VSV-G was cloned into pElucEnv. As a negative control, pElucEnv EnvKO was produced. The *env* gene of pElucEnv EnvKO has a stop codon after its 25th codon. The 2.7-kb SalI-BamHI fragment covering the *env* portion of the pElucEnv WT was subcloned into pGEM7zf(+) (Promega) as a target for the mutagenesis. The site-directed mutagenesis used to create the stop codon was performed with a specific primer set (CTCCTTGGGATGTTGTAG ATCTGTAGTCTACA and TGTAGCACTACAGATCTACAACATCCCAAG GAG). The BsiWI-BamHI fragment of the subclone was sequenced and cloned back into pElucEnv WT.

A reporter vector, pTM3hRL, which has a reporter gene, renilla luciferase, under the control of the T7 promoter, was generated from the pTM3luci vector (1) by replacing the firefly luciferase gene with the humanized renilla luciferase gene. The renilla luciferase gene was amplified from pRL-CMV vector (Promega) by PCR with a primer set (CGACTCACTATAGGCTAGCC and GCT CGAGGCGGCCGCTCTAGAATTAC), cloned into pCR4Blunt-TOPO (Invitrogen, Carlsbad, Calif.), and sequenced before cloning. To generate the T7 RNA-polymerase (T7 RNAPol) expression vector, pCMMP T7RNAPoliresGFP, the gene encoding T7 RNA polymerase, was PCR amplified from pVR-T7-1 (1)

and cloned into the pCMMP retrovirus vector (30). In this vector, the T7 RNAPol gene is followed by an internal ribosome entry site and a GFP gene.

**Cells and antibodies.** COS7 cells, 293 cells, and MAGI cells were grown in Dulbecco's modified essential medium (DMEM; Sigma, St. Louis, Mo.) supplemented with 10% fetal bovine serum (FBS) (HyClone Laboratories, Logan, Utah) and penicillin-streptomycin (Gibco-BRL, Rockville, Md.). MAGI cells were grown in DMEM supplemented with 10% FBS, penicillin-streptomycin, Geneticine (0.2 mg/ml), and hygromycin B (0.1 mg/ml) (16). Jurkat cells and H9 cells were grown in RPMI 1640 (Sigma) supplemented with 10% FBS and penicillin-streptomycin. Cells were kept under conditions of 5% CO<sub>2</sub> in a humidified incubator. OKT4 is an anti-CD4 monoclonal antibody and was obtained from Ortho Diagnostic System (Raritan, N.J.). Anti-gp120 polyclonal antibody was obtained from Fitzgerald Industries International, Inc. (Concord, Mass.). The purified anti-gp120 monoclonal antibody prepared from hybridoma 902 was kindly provided by Y. Yokota of the National Institute of Infectious Diseases, Tokyo, Japan. Hybridoma 902 was obtained from Bruce Chesebro through the AIDS Research and Reference Reagent Program, Division of AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health (7, 33). 293CD4 cells were isolated from 293 cells that had been transfected with pMACS 4-IRESII (Miltenyi Biotec GmbH, Bergisch Gladbach, Germany) by use of a MACSelect4.2 system (Miltenyi Biotec GmbH). Serum from a patient infected with HIV-1 was kindly provided by T. H. Lee of Harvard School of Public Health, Boston, Mass.

**Protein analysis.** The WT and mutant HXB2RU3AN constructs were transfected into COS7 cells by the use of Gene Pulsar II (Bio-Rad, Hercules, Calif.). In brief, COS7 cells were suspended in serum-free DMEM and electroporated with 4 µg of proviral DNA at a setting of 250 kV and 950 µF. At 72 h after transfection, cell and virus lysates were prepared for protein analysis. Transfected COS7 cells were collected by scraping and were centrifuged (Allegra 6KR system; Beckman Coulter, Fullerton, Calif.) at 2,000 × g for 10 min. The cell pellets were dissolved in radioimmunoprecipitation assay lysis buffer (0.05 M Tris-Cl [pH 7.2] including 0.15 M NaCl, 1% Triton X-100, 1% sodium deoxycholate, and 0.1% sodium dodecyl sulfate [SDS]) and centrifuged (Himac CS 120fx system; Hitachi, Tokyo, Japan) at 314,000 × g for 45 min at 4°C. The supernatants were subjected to SDS-polyacrylamide gel electrophoresis (PAGE). To analyze viral proteins, the supernatants from culture media were spun (Allegra 6KR system) (2,000 × g for 20 min) to clear cell debris, filtered through 0.45-µm-pore-size filters (Millipore, Bedford, Mass.), and then centrifuged (SW28 rotor; Beckman Coulter) at 4°C for 1.5 h on 3 ml of a 20% sucrose cushion at 113,000 × g. Virus pellets were dissolved in radioimmunoprecipitation assay lysis buffer for protein analysis by SDS-PAGE. Cell and virus lysates were run on an SDS polyacrylamide gel (DRC, Tokyo, Japan) (7.5 to 15% gradient), and proteins were blotted onto Immobilon-P (Millipore) by passive transfer as described previously (21). The immunoblotting procedure was as follows. Before addition of the serum or monoclonal antibodies, the membranes were blocked with 3% bovine serum albumin (Sigma) dissolved in 0.2% Tween 20-PBS at room temperature for 30 min. Enhanced chemiluminescence (Roche Molecular Biochemicals, Mannheim, Germany) and a lumi-imager (Roche) were used to detect the bands after probing with the patient serum or with anti-gp120 polyclonal antibody. Anti-human- and anti-goat-sheep immunoglobulin G-biotin conjugates and streptavidin-horseradish peroxidase were purchased from Amersham Biosciences UK Limited (Buckinghamshire, United Kingdom).

**MAGI cell assay.** MAGI cells were cultured in a 96-well plate for 48 h and then transfected with 0.25 µg of proviral DNA by FuGene6 (Roche Molecular Biochemicals). At 48 h after transfection, the transfected MAGI cells were fixed using PBS containing 0.5% glutaraldehyde and were stained for β-galactosidase as described previously (16). The nuclei in multinucleated cells were counted. Five randomly selected fields were evaluated for each mutant.

**Infection study.** For the infection study, the virus seed was prepared by transfecting 1 µg of the proviral DNA into 10<sup>6</sup> of COS7 cells by FuGene6. At 72 h after transfection, the culture supernatant was filtered through 0.45-µm-pore-size filters (Millipore) and the p24 amount was determined using a p24 enzyme-linked immunosorbent assay (ELISA). Jurkat cells were infected with each virus adjusted by the p24 amount (10 ng of p24 per 10<sup>6</sup> cells to be infected). The infection was monitored by measuring the p24 amount of the culture supernatant at the indicated time point after infection. A p24 ELISA was performed using a p24 RETRO-TEK ELISA kit (ZeptoMetrix, Buffalo, N.Y.).

**Flow cytometry.** At 48 h after transfection by FuGene6, the COS7 cells that had been transfected with each Env expression vector (pElucEnv WT, MSD GpA, MSD GpAG83I, MSD VSV-G, or EnvKO) were stained with the 902 monoclonal antibody for 1 h at 4°C (10 µg/ml in PBS with 2% FBS), incubated with biotin-XX goat anti-mouse immunoglobulin G (Molecular Probes, Eugene, Oreg.) for 30 min at 4°C, and then treated with streptavidin Alexa Fluor 555

(Molecular Probes) for 30 min at 4°C and finally fixed with 1% paraformaldehyde in PBS. Cells were suspended in PBS with 2% FBS and analyzed with Becton Dickinson FACSCalibur and CellQuest software (BD Biosciences Immunocytometry Systems, San Jose, Calif.). A double gate was defined by forward versus side scatter and by the amount of GFP (FL-1). A total of 10,000 events within this gate were collected for analysis.

**CD4-binding assay.** To evaluate the CD4-binding capacity of each MSD-mutant Env, cell lysates prepared from the COS7 cells transfected with each Env expression vector (pElucEnv WT, MSD GpA, MSD GpAG83I, or MSD VSV-G) and a CD4 expression vector by FuGene6 were prepared. At 72 h after transfection, the transfected COS7 cells were suspended in 250 µl of a binding buffer, 0.5% Triton X-100 in PBS, supplied with the Complete protease-inhibitor cocktail (Roche Molecular Biochemicals). Aliquots of 50 µl from each lysate were mixed together, and 2 µl of OKT-4 and 5 µl of protein G Magnetic Beads (New England BioLabs, Beverly, Mass.) were added. After a 1-h incubation at 4°C, the beads were washed twice with 200 µl of binding buffer and subjected to SDS-PAGE. The gp120 was detected using an immunoblotting analysis (see above).

**Dye-transfer assay.** The COS7 cells were transfected with each Env expression vector (pElucEnv WT, MSD GpA, MSD GpAG83I, EnvKO, or MSD VSV-G) by FuGene6 and cocultured with H9 cells at 48 h after transfection. H9 cells used for this assay were loaded with CellTracker CM-DII (at a final concentration of 10 µM) and Calcein blue, AM (Molecular Probes) (at a final concentration of 40 µM), precultured for 3 h, and then washed with culture medium. At 2 h after coculture, the cells were fixed with 4% paraformaldehyde-PBS and analyzed using a Zeiss LSM510META microscope. Cells showing green, red, or blue fluorescence were counted in each of five randomly selected fields at magnification of ×200.

**T7 RNAPol transfer assay.** The COS7 cells were transfected with each Env expression vector (pElucEnv WT, MSD GpA, MSD GpAG83I, MSD VSV-G, or EnvKO) together with pTM3hRL (see above) by the use of FuGene6. At 48 h after transfection, the transfected COS7 cells were cocultured with the 293CD4 cells that had been transfected with the T7 RNAPol expression vector, pCMMP T7RNAPoliresGFP (the ratio of cells was 1:1). At 12 h after coculture, the cells were lysed and the firefly luciferase activities, derived from the Env expression vector, and renilla luciferase activities, activated by the T7 RNAPol transferred from 293CD4 cells through the generated fusion pores, were determined using a Dual-Glo luciferase reporter assay system (Promega).

## RESULTS

Gly<sup>sub</sup> mutations did not affect the fusion activity of gp41 in MAGI cells. The amino acid sequences of the predicted MSD of HXB2, the consensus sequences of HIV-1 subtypes, and sequences of certain other membrane proteins are shown in Fig. 1A. Although the MSDs of CD22 and CD4 could replace the MSD of gp41 without affecting its function (41, 45), the lengths of the predicted MSDs differ and there is no apparent sequence homology among them except for the presence of several glycine residues. The amino acid sequence of the MSD of gp41 is well conserved among different clades, and it contains the GXXXG motif. Except for the MSD in CD22, all MSDs listed contain the GXXXG motif. To evaluate the relative contributions that the three glycine residues in the gp41 MSD made to fusion activity, each glycine residue was changed to an alanine or a leucine residue. Additionally, all three glycine residues were changed to alanine residues in the 3A mutant. The amino acid sequences and nomenclatures of Gly<sup>sub</sup> mutants are shown in Fig. 1B. Alanine and leucine residues were chosen because they are commonly found in the transmembrane region of membrane proteins (39). In a previous study, substituting alanine or leucine residues for glycine residues in VSV-G produced membrane-fusion-incompetent proteins (8).

The protein profile of Gly<sup>sub</sup> mutants was examined by immunoblotting analysis of both cell and virus lysates derived from COS7 cells transfected with the proviral DNAs. Similar

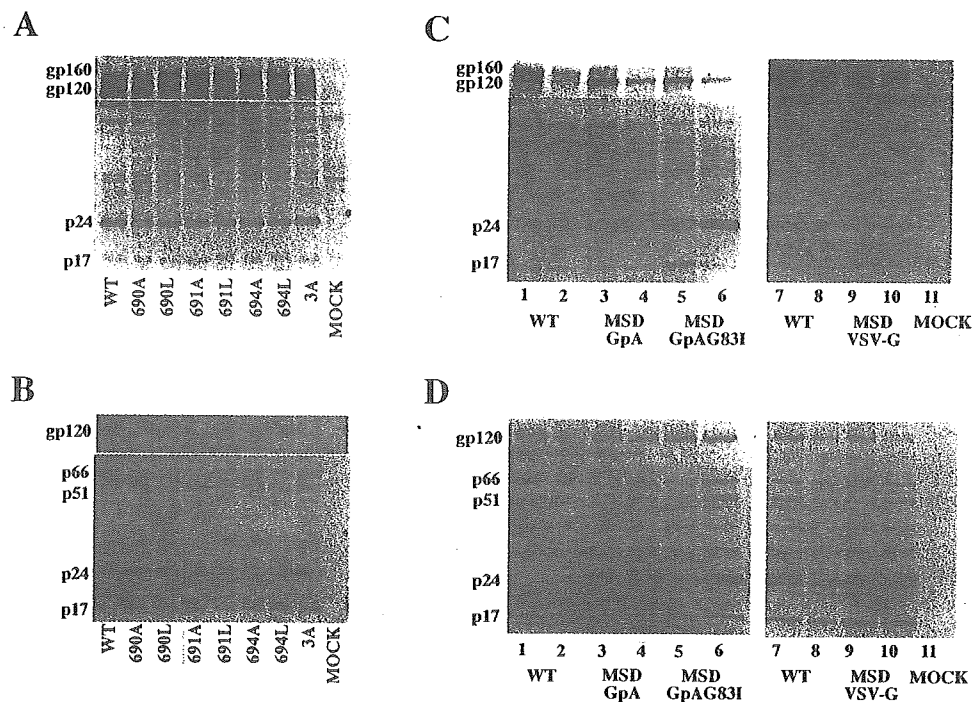


FIG. 2. Protein profiles of cell and virus lysates of the WT and mutants. The cell and virus lysates for glycine substitution (A and B) and MSD replacement mutants (C and D) were prepared from COS7 transfected with proviral DNA. The Env proteins (gp160, gp120) were detected with anti-gp120 polyclonal antibody. Gag (p24, p17) and Pol (p66, p51) were detected using serum from an individual infected with HIV-1. The name of the provirus DNA is indicated below, and the bands corresponding to gp160, gp120, p66, p51, p24, and p17 are shown by arrows. (C and D) For MSD replacement mutants, two different amounts of lysates were loaded for each mutant. The second lane of each sample received an amount of the lysate that was half of the amount used for the preceding lane.

protein profiles were observed for all Gly<sup>sub</sup> mutants and the WT in both cell and virus lysates (Fig. 2A and B). For all Gly<sup>sub</sup> mutants, both gp160 and gp120 were detected in cell lysates and gp120 was detected in virus lysates. We did not observe prominent changes in Env processing, as in the case of the MSD replacement (MSD<sup>cep</sup>) mutants (see below). In virus lysates, similar levels of Env relative to virus Gag/Pol products (p17, p24, p51, and p66) were detected in each sample (Fig. 2B). These data suggested that all the Env mutants were expressed, processed, transported to the plasma membrane, and incorporated onto the virions. Fusion activities of the Gly<sup>sub</sup> mutants were evaluated by a MAGI cell assay using an arbitrary fusion index that reflects both the number of syncytia and the number of nuclei within a syncytium. As shown in Fig. 3, the fusion index of all Gly<sup>sub</sup> mutants was comparable with that of the WT. Notably, the fusion activity of the 3A mutant was also well maintained (Fig. 3). These data suggested that glycine residues in the MSD of gp41 were not necessary for membrane fusion in MAGI cells.

**Gly<sup>sub</sup> mutants retained the replication capacity in Jurkat cells.** We next evaluated the effect of glycine substitution on viral replication in T-cell lines. Viruses collected from the culture supernatants of COS7 cells transfected with the proviral DNAs were used to infect Jurkat or H9 cells. Replication of mutant viruses was monitored by measuring the amount of p24 released into the culture medium. A representative result is shown in Fig. 4. All Gly<sup>sub</sup> mutants were replication competent and showed replication kinetics similar to those of the WT

virus in Jurkat cells. Similar results were observed in H9 cells (data not shown). These results were consistent with the results of fusion assay in MAGI cells (Fig. 3) and suggested that the contribution of glycine residues in MSD to the life cycle of HIV-1 was relatively small.

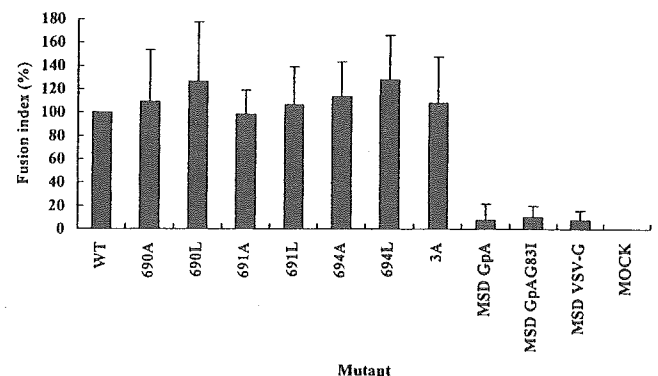


FIG. 3. Relative fusion activity of MSD mutants in MAGI cell assay. Fusion activity of the WT and MSD mutants was expressed using a fusion index (fusion index =  $2x + y$ , where  $x$  is the number of multinucleated cells [number of nuclei  $\geq 5$  in five visual fields] and  $y$  is the number of multinucleated cells [number of nuclei  $< 5$  in five visual fields]). The fusion index was defined to reflect the number of nuclei in multinucleated cells. Fusion activities for each mutant are shown after normalization to that of the WT (the WT activity was set at 100%). Similar results were obtained in three independent experiments.

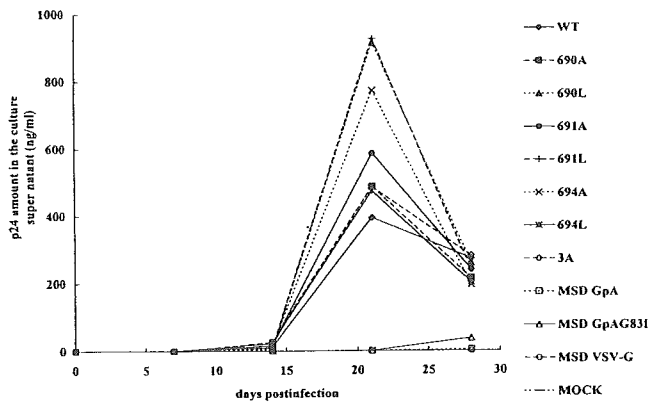


FIG. 4. Replication profiles of the WT and MSD mutants in Jurkat cells. Viral stocks of the WT and each mutant were prepared from culture supernatant of COS7 cells transfected with proviral DNA, and Jurkat cells were infected with each virus adjusted by the amount of p24. Replication was monitored by measuring the amount of p24 in the culture supernatant at specific time points following infection. A representative result of three independent experiments is shown.

The protein profile of MSD<sup>rep</sup> mutants was altered. Our data for Gly<sup>sub</sup> mutants suggested that the MSD of gp41 might have a rather high tolerance for mutation. To examine the extent to which the gp41 MSD tolerated more drastic changes in amino acid sequences, the entire MSD of gp41 was replaced with that of VSV-G, GpA, or GpAG83I. The sequence of each MSD is shown in Fig. 1B. Both GpA and VSV-G contain the GXXXG motif within their MSDs. The former takes a dimer, and the latter forms a trimer. The GpAG83I is a GpA mutant that lacks homodimerization activity owing to the substitution of the 83rd glycine residue (the numbering is according to the system used for GpA study) with an isoleucine residue. This GpAG83I was included to overcome the potential negative effect of the homodimerization activity of GpA (34).

The expression of viral gene products from MSD<sup>rep</sup> mutants was examined by immunoblotting analysis using cell and virus lysates derived from COS7 cells transfected with the proviral DNAs. Two different amounts of each lysate type (cell and virus) were loaded for each construct to achieve semiquantitative evaluation of the band intensity. The second lane of each sample received an amount of the lysate that was half that of the preceding lane. In cell lysates, the reproducible change in the processing pattern of Env was noted for MSD<sup>rep</sup> mutants. All MSD<sup>rep</sup> mutants showed more signal for gp120 than for gp160 (Fig. 2C). This was most prominent in GpA and GpAG83I mutants, followed by the VSV-G mutant (Fig. 2C). The sum intensity of gp160/gp120 normalized by the p24 amount was approximately same for both MSD<sup>rep</sup> mutants and the WT.

This semiquantitative analysis was also performed for virus lysates (Fig. 2D). In virus lysates, the intensity for gp120 relative to p24 was stronger for GpA and GpAG83I mutants (compare lane 1 with lane 4 or 5 in Fig. 2D). Compared with the WT, the VSV-G replacement mutant showed relatively more p24 for an approximately equivalent amount of gp120 (compare lanes 7 and 9 in Fig. 2D). This may indicate that the VSV-G mutant incorporated Env less efficiently than the WT or that Env was shed by the VSV-G mutant. Thus, there were

some alterations in expression, processing, and transport to the plasma membrane and in the incorporation onto the virions in Env of MSD<sup>rep</sup> mutants in the provirus context. Next we evaluated the function of Env of MSD<sup>rep</sup> mutants.

Fusion activity of MSD<sup>rep</sup> mutants was significantly decreased. The MAGI cell assay was used to examine the fusion activity of MSD<sup>rep</sup> mutants. Compared with the WT, far fewer multinucleated cells were observed in MSD<sup>rep</sup> mutants. Although syncytia were occasionally observed in an MSD<sup>rep</sup> mutant, each syncytium observed contained fewer nuclei than any given syncytium observed in the WT. Thus, fusion indices of MSD<sup>rep</sup> mutants were very low compared with those of the WT (Fig. 3). These observations suggested the presence of defects in the membrane fusion steps themselves, because changes in the expression level and the processing of Env in cell lysates (Fig. 2C) alone did not seem to be able to account for the observed severe defect in fusion. These defects were partially compensated for when we used a HeLa-CD4-derived cell line that had a higher level of CD4. In this cell line, the efficiency of syncytia formation and the number of nuclei in a syncytium were increased (data not shown). Consistent with the defect in fusion activity, MSD<sup>rep</sup> mutants showed severely impaired replication capacity in Jurkat cells (Fig. 4) and in H9 cells (data not shown). We tried to determine the point at which membrane fusion ceased in MSD<sup>rep</sup> mutants by using the Env expression vector.

The cell surface expression and CD4-binding capacity of Env for MSD<sup>rep</sup> mutants were similar to those for WT. We constructed the Env expression vectors to analyze the steps of membrane fusion processes of MSD<sup>rep</sup> mutants (Fig. 5A). First, we examined the cell surface Env expression of MSD<sup>rep</sup> mutants by transfecting each Env expression vector into COS7 cells and then performing a flow cytometric analysis using an anti-gp120 monoclonal antibody. As the transfected cells expressed GFP derived from the expression vector, cells were gated for GFP first. A representative result is shown in Fig. 5B. The fluorescence intensities of Env in MSD<sup>rep</sup> mutants and the WT were nearly equal. The expression of Env was also verified by Western blotting analysis (data not shown).

Next, we tested the CD4-binding capacity of the Env of MSD<sup>rep</sup> mutants, because mutations in MSDs potentially affect the conformation of Env. We immunoprecipitated CD4 by using anti-CD4 antibody from a mixture of cell lysates prepared from COS7 cells that had been transfected with either the CD4 expression vector or the Env expression vector. The gp120 that coimmunoprecipitated with the CD4 was detected using immunoblotting analysis. The gp120 coimmunoprecipitated at similar efficiencies in MSD<sup>rep</sup> mutants and the WT (Fig. 5C). These data suggested that the MSD replacement did not affect the CD4-binding capacity of the mutated Envs. Thus, the membrane fusion defect of Envs of MSD<sup>rep</sup> mutants was determined to be in the post-CD4-binding steps.

Examination of the hemifusion and fusion pore formation steps in MSD<sup>rep</sup> mutants. We used the dye-transfer assay to examine whether Envs of MSD<sup>rep</sup> mutants were able to induce lipid mixing and fusion pore formation between Env- and receptor-expressing cells. In this assay, the COS7 cells transfected with a recombinant Env expression vector bearing the EGFP-firefly luciferase hybrid gene (Fig. 5A) were cocultured with T cells whose cell membrane and cytoplasm were labeled



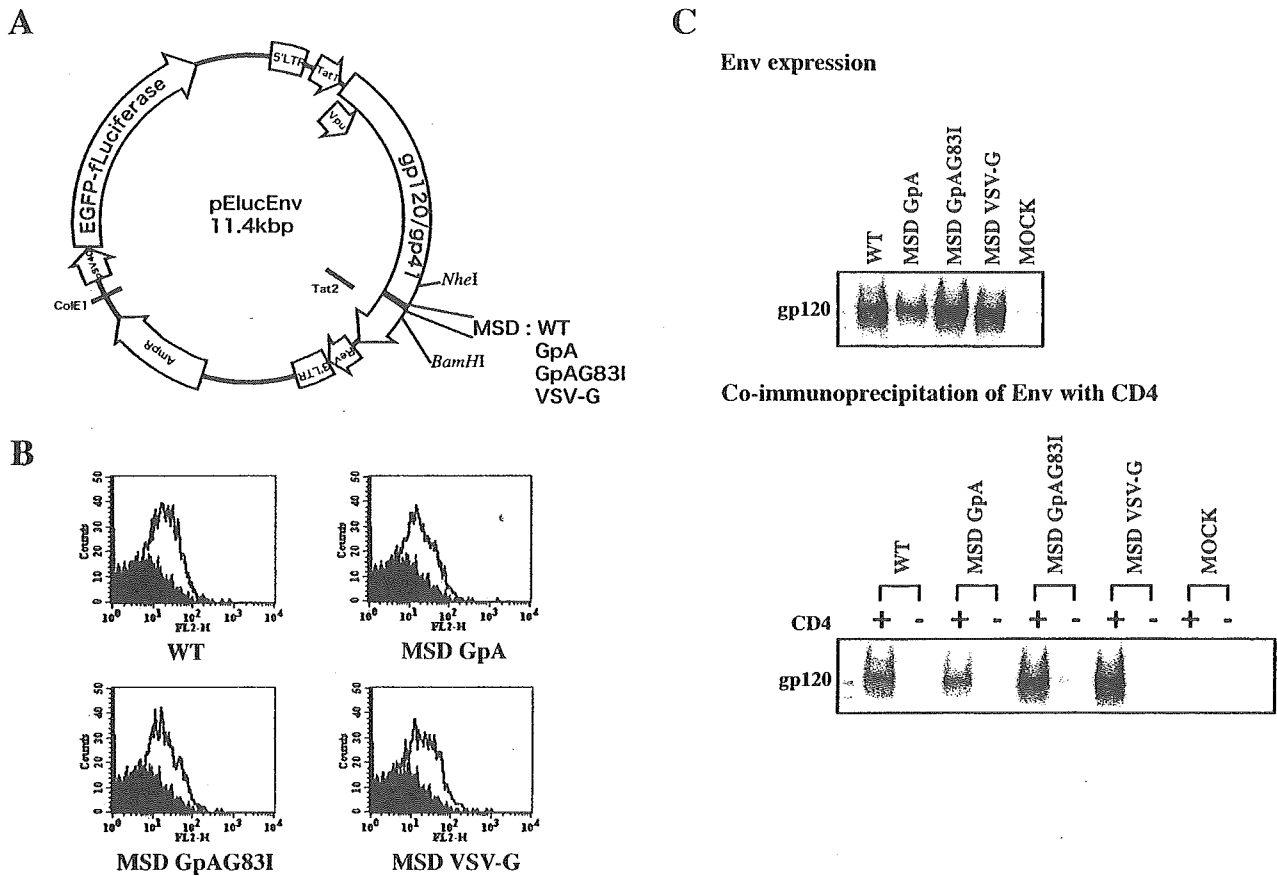


FIG. 5. Evaluation of the cell surface expression and the CD4-binding capacity of Env of MSD mutants. (A) The map of the Env expression vector pElucEnv. pElucEnv supports the expression of HIV-1 *env* (gp120/gp41) and the gene of the EGFP-firefly luciferase (EGFP-luciferase) hybrid protein from two separate promoters. The *NheI*-*BamHI* fragments of pSP65HXB2RU3ΔN WT, MSD GpA, MSD GpAG83I, or MSD VSV-G were cloned into pElucEnv. 5'LTR, 5' long terminal repeat of HIV-1; 3'LTR, 3' long terminal repeat of HIV-1. *tat*, *vpu*, and *rev* represent *tat*, *vpu*, and *rev* of HIV-1. Amp<sup>r</sup>, beta-lactamase; SV40, simian virus 40 late promoter; ColE1, ColE1 replication origin; *NheI* and *BamHI*, restriction sites used for cloning of the mutated *env*. COS7 cells were transfected with each pElucEnv construct and subjected to flow cytometric analysis as described in the Materials and Methods section. The signal for each Env is shown with a gray line. The filled area depicts the signal obtained for the control vector, EnvKO. (C) CD4-binding capacity of Env of MSD mutants. The cell lysates prepared from the COS7 cells that had been transfected with the pElucEnv construct or the CD4 expression vector were mixed together, and CD4 was immunoprecipitated using OKT-4. The amount of gp120 coimmunoprecipitated with CD4 was evaluated. The upper panel shows the expression level of each Env mutant. Coimmunoprecipitated gp120 was detected using anti-gp120 polyclonal antibodies (shown in the lower panel).

with CM-DiI (red) and Calcein blue, AM (blue), respectively. When the lipids from the two cells mix, a transfer of red dye should be observed. If fusion pores are formed, fused cells whose cytoplasm bears both blue and green fluorescence should appear. The results are shown in Fig. 6. At 2 h after coculture, cells having green, red, and blue fluorescence were observed in the WT (Fig. 6A). Similar cells were also observed in MSD<sup>TEP</sup> mutants but were fewer in number than those observed in the WT. We determined the frequency of lipid or cytoplasmic mixing 2 h after coculture by counting Env-expressing cells (green cells) whose membrane or cytoplasm was stained with CM-DiI or Calcein blue, AM, respectively. In MSD<sup>TEP</sup> mutants, many Env-expressing COS7 cells were surrounded by T cells and showed no evidence of dye transfer. The relative frequency of dye transfer was determined to be less than half that of the WT (Fig. 6B). In this assay, we did not observe significant discrepancies in the frequencies with which the two dyes were transferred; that is to say, cells arrested at

the hemifusion stage were not detected. These data suggested that Env of MSD<sup>TEP</sup> mutants retained the capacity for fusion but at efficiencies that were much lower than those of the WT and that failed to produce large syncytia. The decrease in transfer of CM-DiI indicated that MSD<sup>TEP</sup> mutants had a defect that affected the fusion process prior to hemifusion. The smaller size of the syncytia observed in MSD<sup>TEP</sup> mutants may suggest the presence of other defects after hemifusion, possibly an enlargement of fusion pore.

We performed an additional quantitative assessment of fusion pore formation by the use of the T7 RNAPol transfer assay. In this assay, COS7 cells were first transfected with an Env expression vector coupled with a reporter plasmid, pTM3hRL, which encodes T7 promoter-driven renilla luciferase, and then were cocultured with 293CD4 cells that had been transfected with a T7 RNAPol expression plasmid. When the COS7 and 293CD4 cells fused, the T7 RNAPol in the 293CD4 cells transferred into the COS7 cells and drove ex-

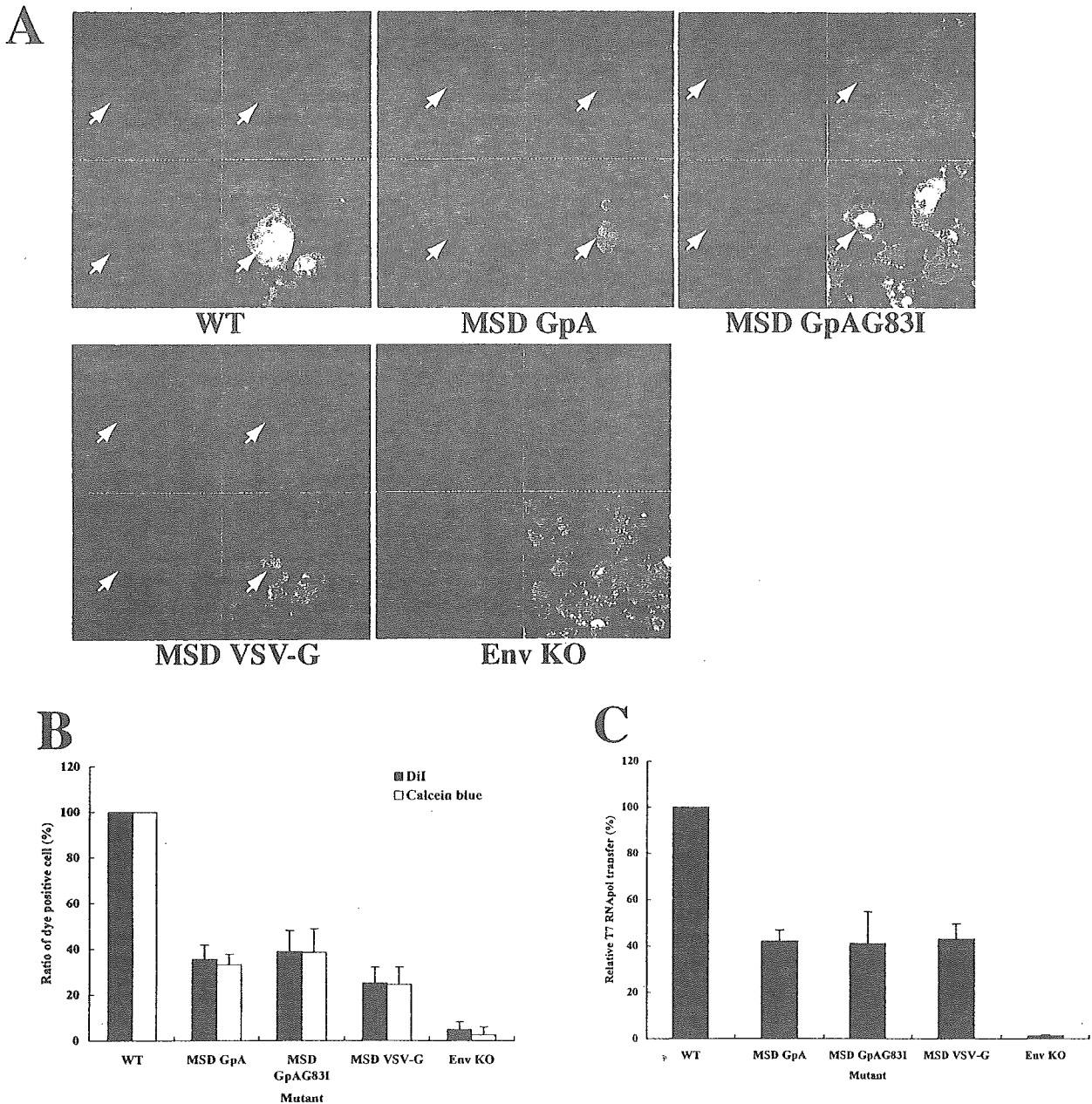


FIG. 6. Cell-cell fusion analysis. The results of a dye-transfer assay using a coculture system and COS and T cells are shown. The COS7 cells transiently transfected with each pElucEnv construct were cocultured with H9 cells that had been loaded with CellTracker CM-DiI (pseudo-red) and Calcein blue, AM (pseudo-blue). Pictures taken at 2 h after coculture are shown in panel A. Typical fused cells having green, red, or blue fluorescence are indicated by arrows. Original magnification,  $\times 200$ . (B) Relative dye-transfer frequencies of MSD<sup>rcp</sup> mutants. The number of Env-expressing cells (green cells) whose membranes or cytoplasm are labeled with CM-DiI (red) or Calcein blue, AM (blue), respectively, was counted in five, randomly selected fields. The value was established by setting WT at 100%. The averages of the results from four experiments are shown. (C) Results of T7 RNApol-transfer assay to determine fusion pore formation of MSD mutants. The COS7 cells were transfected with each pElucEnv construct and the T7 RNApol-responsive reporter plasmid, pTM3hRL. At 48 h after transfection, the transfected COS7 cells were cocultured with the 293CD4 cells that had been transfected with pCMMP T7RNApoliresGFP. At 12 h after coculture, the cells were lysed and firefly luciferase and renilla luciferase activity levels were measured. Pore formation efficiencies were calculated by comparing the induced renilla luciferase activity to the firefly luciferase activity. A representative result of four independent experiments is shown.

pression of renilla luciferase. At 12 h after coculture, renilla luciferase activities were measured. The value was normalized according to the efficiency of transfection measured by firefly luciferase activities derived from the Env expression vector. Representative results are shown in Fig. 6C. Normalized re-

nilla luciferase activities of MSD GpA, MSD GpAG83I, MSD VSV-G, and EnvKO were, respectively, 42, 41, 43, and 1.3% of that of the WT. Thus, the transfer of T7 RNApol observed for cells that were expressing MSD<sup>rcp</sup> Env was less efficient than that of the WT, a finding consistent with that from the dye-

transfer assay. We also tested a cleavage site mutant of Env (R511S) (3), which cannot produce an active fusion peptide and is unable to induce membrane fusion in this assay. The value was 2.6%. Therefore, the value obtained for MSD<sup>rep</sup> mutants should reflect the efficiency of de novo pore formation. The value of around 40% of that of the WT obtained for MSD<sup>rep</sup> mutants may appear high compared with the severe defects in the MAGI assay (Fig. 3) and replication assay (Fig. 4). This could be the result of differences in the two assay systems or it might suggest the presence of additional defects in membrane fusion in MSD<sup>rep</sup> mutants.

## DISCUSSION

The involvement of the MSD in membrane fusion has been studied for several different viruses, including HIV-1 (8, 14, 15, 22, 23, 28, 35, 36, 38, 42). However, the precise molecular mechanism of Env-mediated membrane fusion is not known, and the role of MSD, including whether the specific amino acid sequences in MSD are required for the membrane fusion process, is not well understood. Examination of the presently available HIV-1 sequence database reveals a high conservation of MSD sequences for gp41 (17). There also are several well-conserved glycine residues that form the GXXXG motif, a motif found in the helix-helix interface of many membrane-spanning  $\alpha$ -helices (24, 26, 29). Interestingly, the MSDs of CD4 and CD22, which can replace the MSD of gp41 without affecting its fusion activity, also contain several glycine residues. Building from these findings, we investigated whether the specific sequence of the gp41 MSD is required for competent membrane fusion. We did this by mutating the conserved glycine residues in the gp41 MSD to alanine or leucine residues. A similar mutation of glycine introduced into VSV-G has a detrimental effect on membrane fusion (8). Contrary to our expectations, mutation of glycine residues resulted in fusion-competent gp41 (Fig. 3). Even replacing all three glycine residues in gp41 MSD with alanine residues did not affect fusion activity (Fig. 3). This suggests that, *in vitro*, the gp41 MSD has a rather high tolerance for mutation. The conservation of glycine residues observed in field isolates may indicate that HIV-1 requires glycine residues for *in vivo* infection of cells whose properties are different from those seen with *in vitro* T-cell lines. This may also imply that the potential mechanisms involved in membrane fusion differ for gp41 and VSV-G. It may be that the contribution of the proposed kink in the MSD, induced by the presence of a glycine residue during VSV-G-mediated membrane fusion, is not as critical to the membrane fusion process in gp41 as it is for the same process in VSV-G (8). To further verify the role of GXXXG motif for a potential helix-helix association motif, future analyses may require simultaneous substitution of all glycines with other bulky amino acid residues.

Proceeding on the basis of our results obtained with Gly<sup>sub</sup> mutants and previous reports of replication-competent HIV-1 having heterologous MSD in place of the natural gp41 MSD, we next replaced the entire gp41 MSD with MSDs of heterologous membrane proteins (GpA or VSV-G). Membrane fusion activity was severely impaired in both the GpA and VSV-G replacement mutants (Fig. 3). Because GpA is known to form a dimer through the GXXXG motif, the dimerization

activity may interfere with the proper trimerization of gp41 and cause the observed defect. However, the observed defect was not rescued by introducing a G83I mutation within the MSD of GpA (Fig. 3), a mutation that has been shown to knock out the dimerization ability (34). Therefore, the defect cannot be explained simply by noting that replacing the MSD interfered with gp41's proper trimerization. A similar defect in membrane fusion observed in our VSV-G-replacement mutant may further support this argument, because VSV-G, like gp41, is reported to form a trimer. At this time, however, we cannot completely rule out the potential defect in proper trimerization of Env in the MSD replacement mutants.

It is worth mentioning that we observed more gp120 than gp160 reproducibly in MSD<sup>rep</sup> mutants, especially for GpA-replacement mutants in the transfected cells. This may suggest that there are some subtle conformational changes in Env for the GpA-replacement mutants that allow for more efficient processing by the furin-like protease. It also suggests that the mutation may alter the length of time needed for the mutant envelope proteins to travel through the *trans* Golgi network. These possibilities are consistent with a hypothesis that mutation within the MSD may affect the intracellular trafficking of the envelope protein. We cannot rule out this hypothesis.

In the provirus context, a smaller amount of Env was detected in virus lysates for VSV-G replacement mutant (Fig. 2D). This defect, in addition to the defect in fusion, may contribute to the VSV-G replacement mutant's inability to replicate (Fig. 4). The mechanism of this defect is not known at present. The mutation within the MSD, however, may affect the structure of gp41 and that, in turn, may result in a less efficient interaction between Gag and Env during assembly or a less stable association between gp41 and gp120 in this replacement mutant.

To analyze the effect of the MSD mutation in membrane fusion independent of other viral structural proteins, we used the Env expression vector. We found no significant difference in the steady-state amount of envelope proteins expressed on the cell surface and the CD4-binding efficiency (Fig. 5B). Therefore, the drastic changes in either trafficking or conformation cannot explain the observed prominent defect in fusion activity of GpA or VSV-G replacement mutants. One of the potential defects of membrane fusion in our MSD<sup>rep</sup> mutants should reside in the post-CD4 binding step(s) of membrane fusion processes. In MSD<sup>rep</sup> mutants, there may be a difference in the interactions between Env and the chemokine receptors or in the conformational changes induced after such interactions.

The exact defective steps in the membrane fusion mechanism of MSD<sup>rep</sup> mutants remain to be determined. There should be a defect before the lipid dye mixing step, because we observed decreased efficiency of lipid dye transfer in MSD<sup>rep</sup> mutants (Fig. 6B). We did not observe the discrepancy between lipid dye (CM-DiI) and cytoplasmic dye (Calecein blue, AM) transfer; therefore, there was no arrest in the hemifusion step. Recently, Lin and coworkers reported that truncating the MSD of simian immunodeficiency virus (SIV) decreased fusion activity (18). As with our mutants, their SIV Env mutants exhibited an overall decrease in lipid dye transfer as well as in cytoplasmic dye transfer. Theirs, however, showed a further downstream defect, namely, an enlargement of the fusion pore

induced by mutated Env. This was suggested by the difference in the efficiency of transfer between low-molecular-weight molecules (Calcein blue, AM) and larger molecules (enhanced blue fluorescent protein). We did not observe this difference in our assay. They also observed the gap between the lipid dye transfer and luciferase transfer assays. In our mutants, the decrease in the lipid dye transfer correlated well with the decrease in the transfer of T7 RNAPol assay (Fig. 6B and C). The discrepancy between the studies might arise from the differences between the two viruses (SIV and HIV) or from the nature of the mutations introduced (truncation versus replacement).

We observed a pore formation efficiency of approximately 40% for the MSD<sup>rep</sup> mutants compared with that observed with the WT despite their prominent fusion defect observed in MAGI cells. Because the two assays described in Fig. 3 and 6 employed different cell types and expression systems, it may be difficult to compare these two assays directly. There is the possibility that the discrepancy between the assays was artificially enhanced by the differences in the two assay systems. For example, the levels of expressed Env and of CD4 were higher in the T7 RNAPol transfer assay than in the MAGI cell assay (data not shown). When a higher CD4-expressing cell line was used in the fusion assay instead of the MAGI cells, while maintaining with the provirus constructs, the efficiency of fusion was increased (data not shown). When the provirus constructs were used instead of the Env expression vectors in the T7 RNAPol transfer assay, the values for pore formation were slightly decreased. The individual values of GpA, GpAG83I, and VSV-G were 39, 27, and 20%, respectively. The value for the mock treatment was about 1%. Thus, the observed gap could be accounted for in part by the difference in assay systems. However, very inefficient formation of the large syncytia in MSD<sup>rep</sup> mutants might indeed suggest the presence of the defect in the steps after the formation of initial fusion pores.

Our data on MSD<sup>rep</sup> mutants clearly demonstrate that not all heterologous MSDs can replace the gp41 MSD without affecting its function. The data indicate that the expression or retention of gp41 on the lipid bilayer and the fusion activity itself are rather independent functions. Our results also show that the mere maintenance of the GXXXG motif is not sufficient for fusion activity. It seems likely that a context-dependent arrangement of the glycine and other amino acid residues within the MSD is critical for functional integrity.

In this study, we have provided evidence that the MSD of gp41 affects the biogenesis of Env and also plays a critical role in membrane fusion for HIV-1. As our results for the Gly<sup>sub</sup> and MSD<sup>rep</sup> mutants suggest, the gp41 MSD shows a rather high tolerance for mutation but does require the MSD to have some specific sequences—or structures generated by them—for its proper function. One possible scenario is that the gp41 MSD may interact with some lipid or protein components during membrane fusion. Such an interaction between Env and lipid or protein moieties has been reported for the Semliki Forest virus and the influenza virus, respectively (2, 19). The gp41 MSDs might interact among themselves, with fusion peptides, or with the MSDs of other host proteins, such as CD4 or chemokine receptors, that are thought to come within close proximity of one another during the membrane fusion process. A failure to properly interact because of mutations in the gp41

MSD may affect the fusion process. To address these issues, additional systematic mutagenesis studies are needed to determine the critical residues of the MSD. Such studies should shed light on the molecular mechanisms of membrane fusion. Elucidating the precise molecular mechanism of membrane fusion—and the role of the gp41 MSD in it—may provide another target for a molecular intervention in HIV-1 infection.

#### ACKNOWLEDGMENTS

This study was supported by the Health and Labour Sciences Research Grants from Japanese Ministry of Health, Labor, and Welfare. We thank A. M. Menting for assistance in manuscript preparation.

#### REFERENCES

- Aoki, Y., H. Aizaki, T. Shimoike, H. Tani, K. Ishii, I. Saito, Y. Matsuura, and T. Miyamura. 1998. A human liver cell line exhibits efficient translation of HCV RNAs produced by a recombinant adenovirus expressing T7 RNA polymerase. *Virology* 250:140–150.
- Armstrong, R. T., A. S. Kushnir, and J. M. White. 2000. The transmembrane domain of influenza hemagglutinin exhibits a stringent length requirement to support the hemifusion to fusion transition. *J. Cell Biol.* 151:425–437.
- Bosch, V., and M. Pawlita. 1990. Mutational analysis of the human immunodeficiency virus type 1 *env* gene product proteolytic cleavage site. *J. Virol.* 64:2337–2344.
- Bullough, P. A., F. M. Hughson, J. J. Skehel, and D. C. Wiley. 1994. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* 371:37–43.
- Chan, D. C., D. Fass, J. M. Berger, and P. S. Kim. 1997. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* 89:263–273.
- Chen, J., S. A. Wharton, W. Weissenhorn, L. J. Calder, F. M. Hughson, J. J. Skehel, and D. C. Wiley. 1995. A soluble domain of the membrane-anchoring chain of influenza virus hemagglutinin (HA2) folds in *Escherichia coli* into the low-pH-induced conformation. *Proc. Natl. Acad. Sci. USA* 92:12205–12209.
- Chesebro, B., and K. Wehrly. 1988. Development of a sensitive quantitative focal assay for human immunodeficiency virus infectivity. *J. Virol.* 62:3779–3788.
- Cleverley, D. Z., and J. Lenard. 1998. The transmembrane domain in viral fusion: essential role for a conserved glycine residue in vesicular stomatitis virus G protein. *Proc. Natl. Acad. Sci. USA* 95:3425–3430.
- Dubay, J. W., S. J. Roberts, B. H. Hahn, and E. Hunter. 1992. Truncation of the human immunodeficiency virus type 1 transmembrane glycoprotein cytoplasmic domain blocks virus infectivity. *J. Virol.* 66:6616–6625.
- Eisenberg, D., and M. Wesson. 1990. The most highly amphiphilic alpha-helices include two amino acid segments in human immunodeficiency virus glycoprotein 41. *Biopolymers* 29:171–177.
- Fass, D., S. C. Harrison, and P. S. Kim. 1996. Retrovirus envelope domain at 1.7 angstrom resolution. *Nat. Struct. Biol.* 3:465–469.
- Fleming, K. G., and D. M. Engelman. 2001. Specificity in transmembrane helix-helix interactions can define a hierarchy of stability for sequence variants. *Proc. Natl. Acad. Sci. USA* 98:14340–14344.
- Freed, E. O., and M. A. Martin. 1996. Domains of the human immunodeficiency virus type 1 matrix and gp41 cytoplasmic tail required for envelope incorporation into virions. *J. Virol.* 70:341–351.
- Harman, A., H. Browne, and T. Minson. 2002. The transmembrane domain and cytoplasmic tail of herpes simplex virus type 1 glycoprotein H play a role in membrane fusion. *J. Virol.* 76:10708–10716.
- Kemble, G. W., T. Danieli, and J. M. White. 1994. Lipid-anchored influenza hemagglutinin promotes hemifusion, not complete fusion. *Cell* 76:383–391.
- Kimpton, J., and M. Emerman. 1992. Detection of replication-competent and pseudotyped human immunodeficiency virus with a sensitive cell line on the basis of activation of an integrated  $\beta$ -galactosidase gene. *J. Virol.* 66:2232–2239.
- Kuiken, C., B. Foley, B. Hahn, P. Marx, F. McCutchan, J. Mellors, S. Wolinsky, and B. Korber. 2001. HIV Sequence Compendium 2001. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, N.M.
- Lin, X., C. A. Derdeyn, R. Blumenthal, J. West, and E. Hunter. 2003. Progressive truncations C terminal to the membrane-spanning domain of simian immunodeficiency virus Env reduce fusogenicity and increase concentration dependence of Env for fusion. *J. Virol.* 77:7067–7077.
- Lu, Y. E., T. Cassese, and M. Kielian. 1999. The cholesterol requirement for Sindbis virus entry and exit and characterization of a spike protein region involved in cholesterol dependence. *J. Virol.* 73:4272–4278.
- Malashkevich, V. N., B. J. Schneider, M. L. McNally, M. A. Milhollen, J. X. Pang, and P. S. Kim. 1999. Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9-Å resolution. *Proc. Natl. Acad. Sci. USA* 96:2662–2667.