difference in entropy is disregarded in the comparison of the binding energy in this study. ΔG_{MM} is calculated from molecular mechanics (MM) interaction energy:

$$\Delta G_{MM} = \Delta G_{int}^{ele} + \Delta G_{int}^{vdw}$$

where ΔG_{int}^{ele} and ΔG_{int}^{vdw} are electrostatic and van der Waals interaction energies between a ligand and a protein. These energies are computed using the same parameter set with the simulation, and no cutoff is applied for the calculation. Solvation energy ΔG_{sol} can be divided into to the two parts:

$$\Delta G_{sol} = \Delta G_{sol}^{ele} + \Delta G_{sol}^{nonpol}$$

The electrostatic contribution to the solvation free energy (ΔG_{sol}^{ele}) is calculated with the Poisson−Boltzmann method using DelPhi program.[51] The hydrophobic contribution to the solvation free energy (ΔG_{sol}^{nonpol}) is determined as a function of the solvent-accessible surface area.[52]

Results

Conformational Changes of the Protease. To study the mechanism of the drug resistance, we compared the averaged atom coordinates of the wild-type with those of each mutant. The least-squared rigid body superposition indicates that no large conformational alteration appears in the shape of the protease for every mutant, as shown in the left column of Figure 3. The RMSD measurement was executed by using the coordinates of backbone atoms N, Cα, and C in the superimposed structures of the wild-type and each mutant. The RMSD values compared with the wild-type model is 0.8 Å in the D30N, 0.9 Å in the N88D, 1.1 Å in the L90M, and 0.6 Å in the D30N/N88D mutant model, averaged over the whole structure. The detailed analysis of individual residues and the comparison of the local structures, however, provided new understanding on the conformational changes due to the mutation. Although large deviations are seen at the residues of the outside loop region, those residues are not minutely examined because their large fluctuations are commonly observed irrelevant to the mutations in B-factor analysis.[43,53] The most characteristic conformational change is detected in the L90M mutant structure. The L90M mutant structure of our study displays significant backbone deviations at the flap regions (43K-58Q, 43′K-58′Q) and at the 79′P loop (78′G-84′I). The RMSD values are 1.6 Å at the flap of one subunit, 2.0 Å at the opposite flap, and 1.9 Å at the 79′P loop. Those values are very large compared with those of the other mutants (0.5 Å/0.6 Å/0.7 Å in D30N, 0.6 Å/0.7 Å/0.7 Å in N88D, and 0.5 Å/0.6 Å/0.5 Å in D30N/N88D). It should be emphasized that, though the 90th residue is not located at the flap nor at the 79′P loop, L90M affects the conformation at those regions. In the N88D mutant structure, peculiar conformational changes occur at the β sheets consisting of 59Y-75V (RMSD: 1.4 Å) and 59′Y-75′V (RMSD: 1.2 Å). Those regions move far away from the helix region (87R-90L/87′R-90′L). The D30N/N88D mutant structure exhibits the similar conformational changes at the same but much narrower regions of those β sheets: 74T/74′T and its vicinity.

To interpret small and more detailed conformational differences, we compared pairwise Cα−Cα distances in the averaged coordinates of each mutant model with those of the wild-type. Each difference distance is shown in the lower left panel of the map in Figure 4, and the upper right panel shows the error-scaled difference distance.[44,45] Figure 4 also indicates that the L90M model has large conformational changes at the flap region (50I/50′I and its vicinity) and at the loop region at 79′P. The N88D and D30N/N88D models alter the conformations at the β sheets (74′T and its vicinity). In addition, these maps provide more detailed comprehension. The flap conformational alteration in the L90M mutant is owing to the approach of 50I to the triads (25D26T27G/25′D26′T27′G) and the detachment of 50′I from the triads. 79′P, 81′P, and some residues at the same loop are far apart from the residues at the opposite subunit. This loop moves outward and also creates a distance from NFV. The mutants other than L90M exhibit the common structural alteration. For example, the distance between the triad and the flap region is slightly changed. The RMSD from the wild-type at the triads in each subunit is 0.8 Å/1.0 Å (D30N), 0.8 Å/1.2 Å (N88D), and 0.8 Å/1.1 Å (D30N/N88D). These deviations are much larger, compared to the RMSD at the flap region. Hence, characteristic distortion in D30N, N88D, and D30N/N88D is mainly caused by the conformational change of the triad.

The above conformational changes were seen in the last 100 ps of simulation. To ensure that these changes were not just from the local fluctuations of MD simulations, we examined the coordinates for the last 500 ps of simulation. The analysis of the last 500 ps of simulation brought us similar results about the conformational changes. L90M had large deviations at the flap (RMSD: 1.5 Å/1.8 Å) and at the 79′P loop (1.7 Å), compared with wild-type structure. The N88D mutation induced the conformational change at the β sheets around 74T/74′T. In N88D, the RMSD values of the β sheets are 1.4 and 1.2 Å in each subunit. And in D30N/N88D, they are both 0.9 Å. Furthermore, the RMSD values at the triads are 0.8 Å/1.1 Å in D30N, 0.7 Å/1.1 Å in N88D, and 0.8 Å/1.1 Å in D30N/N88D.

Hydrophilic Interactions between NFV and the Protease. In protease−ligand interactions, hydrogen bonds play a crucial role in stabilizing the complex. Hydrogen bonds between the protease and NFV in every model are listed in Table 1. Analysis of the protease−NFV hydrogen bonds suggests that mutation has obviously affected the protease−NFV hydrogen bond network. Only a few residues are responsible for the hydrogen bonds with NFV. Hydrogen bond networks consist of 25D/25′D (catalytic aspartates), 50I/50′I, 30D, 29′D, some water molecules, and NFV. The direct interaction between the carboxyl group of 25′D and the central hydroxyl group of NFV is very frequently seen in every model, whereas the protonated aspartate 25D hardly interacts with NFV, except for the L90M model. At the S2 pocket, the main chain of 30D interacts with the *m*-phenol group of NFV. In the wild-type, the main chain of 30D directly interacts with NFV. In each of the N88D and L90M models, one water molecule links 30D and NFV with a hydrogen bond chain although the distance between the donor and the acceptor atoms is elongated. On the other hand, no hydrogen bond is observed in the mutated D30N and D30N/N88D models. D30N mutation causes the disappearance of the hydrogen bond network at the S2 pocket, because those mutants increase the distances with the *m*-phenol group of NFV and cannot keep any water molecule. At the flap region, one water molecule exists in the neighbor of 50I and/or 50′I and NFV in all models. This water intermediates the intramolecular and intermolecular hydrogen bonds, which results in stabilizing the protease−NFV complex. In the wild-type model, the water links the main chain of 50′I with NFV. Further, a direct hydrogen bond between the sulfur atom of NFV and the main chain of 50I is detected and no water molecule mediates the hydrogen bond from 50I. In contrast, the water mediates the hydrogen bond between NFV and both 50I and 50′I in the D30N model. NFV forms a hydrogen bond network with either 50I or 50′I in the other

TABLE 1: Hydrogen Bond Network in Each Model[a]

		hydrogen bond					
		donor			acceptor		occupancy (%)
wild	catalytic domain	O21	−HOL	(NFV)	OD1	(25′D)	91.0
		O21	−HOL	(NFV)	OD2	(25′D)	95.5
		O	−H2	(WAT208)	O21	(NFV)	60.0
		O	−H1	(WAT208)	OD2	(25′D)	61.0
		N22	−HNM	(NFV)	O	(WAT208)	59.5
	S2 pocket	O38	−HO	(NFV)	O	(30D)	90.0
		N	−H	(30D)	O38	(NFV)	64.5
	flap region	N	−H	(50I)	S74	(NFV)	61.0
		O	−H1	(WAT205)	O17	(NFV)	51.5
		N	−H	(50′I)	O	(WAT205)	61.0
D30N	catalytic domain	O21	−HOL	(NFV)	OD2	(25′D)	100.0
		N22	−HNM	(NFV)	O	(WAT210)	85.0
		O	−H1	(WAT210)	O21	(NFV)	52.5
	flap region	O	−H2	(WAT207)	O17	(NFV)	76.5
		O	−H1	(WAT207)	O25	(NFV)	89.0
		N	−H	(50I)	O	(WAT207)	50.5
		N	−H	(50′I)	O	(WAT207)	93.5
	S2′ pocket	N12	−HNC	(NFV)	O	(WAT4757)	94.5
		N	−H	(29′D)	O	(WAT4757)	85.0
		O	−H2	(WAT4757)	OD1	(29′D)	52.5
N88D	catalytic domain	O21	−HOL	(NFV)	OD2	(25′D)	100.0
	S2 pocket	O	−H1	(WAT3880)	O38	(NFV)	59.0
		N	−H	(30D)	O	(WAT3880)	87.5
		O	−H2	(WAT3880)	O	(30D)	97.5
	flap region	O	−H2	(WAT203)	O17	(NFV)	79.5
		O	−H1	(WAT203)	O25	(NFV)	95.0
		N	−H	(50′I)	O	(WAT203)	100.0
D30N/N88D	catalytic domain	O21	−HOL	(NFV)	OD2	(25′D)	96.5
	flap region	O	−H2	(WAT205)	O25	(NFV)	98.5
		O	−H1	(WAT205)	O17	(NFV)	97.5
		N	−H	(50′I)	O	(WAT205)	96.5
L90M	catalytic domain	O21	−HOL	(NFV)	OD2	(25′D)	100.0
		OD2	−HD2	(25D)	O21	(NFV)	90.5
	S2 pocket	O	−H1	(WAT3836)	O38	(NFV)	73.5
		N	−H	(30D)	O	(WAT3836)	90.0
		O	−H2	(WAT3836)	O	(30D)	99.5
	flap region	O	−H2	(WAT205)	O17	(NFV)	96.5
		O	−H1	(WAT205)	O25	(NFV)	97.5
		N	−H	(50I)	O	(WAT205)	90.5
	S2′ pocket	N12	−HNC	(NFV)	O	(WAT4349)	93.5
		N	−H	(29′D)	O	(WAT4349)	74.5
		O	−H2	(WAT4349)	OD1	(29′D)	74.5

[a] The listed donor and acceptor pairs satisfy the criteria for the hydrogen bond over 50.0% of the time during the last 100 ps of simulation. The nomenclature of the atoms is the same as that of 1OHR.

models (L90M, N88D, and D30N/N88D). Both NFV and the water molecule hardly interact with main chain of 50′I in the L90M mutant and with the main chain of 50I in the N88D and D30N/N88D mutants. Hence, each of the N88D and L90M mutations weakens the hydrogen bond networks around 50I and 50′I, respectively. Interestingly, a water molecule mediating a hydrogen bond with 29′D in the S2 pocket is frequently observed in the simulations of D30N and L90M mutants. Another water molecule mediating an intramolecular hydrogen bond in NFV is located near the catalytic aspartates in both the wild-type and the D30N models (WAT208 and WAT207, respectively). In the wild-type model, this water simultaneously mediates the hydrogen bond between 25′D and NFV.

In addition to the NFV−protease interaction at the active site, the N88D mutation modulates the hydrogen bond network at the surroundings of the 88th residues (Figure 5). In the wild-type, the main chain of 88N has a direct hydrogen bond with the main chain of 29D, and the side chain of 88N interacts with the side chain of 31T. Those hydrogen bonds are also observed in the opposite subunit. Further, one water molecule links 88N with 74T (occupancy is 93.5%) and 88′N with 74′T (75.5%). In each of the D30N and L90M, the 88th residues in both

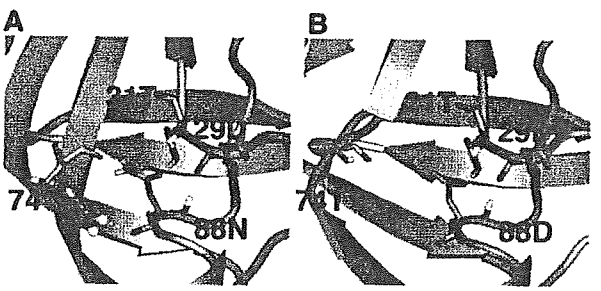


Figure 5. Hydrogen bond network surrounding the 88th residue. A: Wild-type protease structure. B: N88D mutant protease structure.

subunits also have the same hydrogen bond networks as the 29th, 31st, and 74th residues. But in the N88D and D30N/N88D mutant, the water-mediated hydrogen bond does not exist. That is, the N88D mutation induces the disappearance of the hydrogen bonds mediated by water molecules, though the direct hydrogen bonds are retained.

Hydrophobic Interactions between NFV and the Protease. Hydrophobic interactions and van der Waals interactions also contribute to stabilizing the complex. We evaluated the SA

TABLE 2: Buried SA and the Contribution of Hydrophobic and Hydrophilic Residues

	buried SA (Å²)	hydrophobic/hydrophilic[a]
wild	805.9	83.0:17.0
D30N	809.9	82.8:17.2
N88D	841.5	83.1:16.9
D30N/N88D	832.6	86.1:13.9
L90M	765.2	87.1:12.9

[a] Hydrophobic residue: Gly/Ala/Val/Leu/Ile/Met/Pro/Phe/Trp. Hydrophilic residue: Ser/Thr/Asn/Gln/Tyr/Cys/Lys/Arg/His/Asp/Glu.

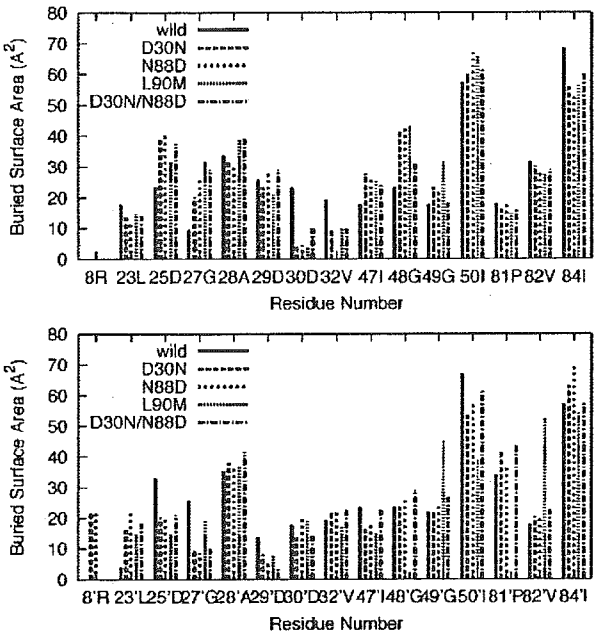


Figure 6. Buried SAs due to each involved residue. The upper graph represents those of one subunit, and the lower graph represents those of the other subunit. L90M shows notable differences at 48'G, 81'P, and 82'V from the other mutants.

buried by the complexation, which is related with the hydrophobicity of the binding cavity and the magnitude of van der Waals contacts between the ligand and the enzyme.[54-56] The results in Table 2 and Figure 6 show that most of the buried residues are hydrophobic ones. The 2-methyl-3-hydroxylbenzamide moiety of NFV has hydrophobic contacts with 23L, 28A, and 84I of the protease at the S2 pocket, and the *tert*-butylcarboxamide moiety has contacts with 32'V and 47'I at the S2' pocket. Hydrophobic interactions are also seen between the dodecahydroisoquinoline ring and 23L, 28A, 81P, 82V, 84I, and 50'I of the protease at the S1 pocket and between the S-phenyl group and 50I, 23'L, 28'A, 81'P, 82'V, and 84'I at the S1' pocket. The stabilization caused by hydrophobic interaction is mainly due to 50I/50'I and 84I/84'I, and secondly 28A/28'A, 48G/48'G, and 81'P. The L90M mutant hardly has hydrophobic contacts with NFV at 48'G and 81'P. These little hydrophobic contributions at 48'G and 81'P result in the serious decrease of the buried surface. The D30N mutation hardly affects the buried SA, whereas the N88D mutation increases the buried SA.

Binding Energy Calculation. Table 3 shows the binding energy between NFV and protease. Each mutation causes a distinctive energetic change from the wild-type. D30N decreases the electrostatic energy greatly. N88D becomes more stable because of the electrostatic contribution by the solvation effect than the wild-type. The D30N/N88D model has both characters of D30N and N88D models. D30N/N88D is less stable than N88D because of the electrostatic contribution and is more stable than D30N by the solvation effect. L90M causes a decrease of the electrostatic energy and a large decrease of the van der Waals energy. These energetic results are compatible with the indices of the resistance level that were estimated from experimental IC90.[8,9]

Discussion

We executed MD simulations to understand the resistance mechanism against NFV about not only active site mutation D30N but also N88D and L90M. The simulations suggest that these mutations affect the protease structures on complexation and the NFV−protease interactions, despite the location of the mutated residues.

D30N is a primary mutation of NFV, which emerges during the treatment with this inhibitor highly specifically. The X-ray crystal structure of the wild-type protease complexed with NFV[13] shows that the *m*-phenol group of NFV interacts with both main chain atoms and the side chain carboxyl group of 30D at the S2 pocket. Accordingly, the D30N mutation has been assumed to make less hydrophilic contacts and causes the disappearance of the hydrogen bond interaction from the 30th residue.

Clemente et al. investigated the D30N mutant protease using a docking study previously.[33] They concluded that the D30N mutant would maintain the hydrogen bond between 30N and NFV but weakens the strength of hydrogen bonding due to the decrease of acid−base interaction. However, their computational model could move only NFV and the side chain of 30N in the vacuum condition without any water molecules. That is, their docking simulations did not consider the contributions of the movement of the residues other than 30N nor water solvation effects. MD simulation is useful to incorporate these contributions and to provide more detailed information. The proton or water-mediated hydrogen bonds are observed between the *m*-phenol group of NFV and the main chain of 30D in the wild-type model and the other models having a sequence of 30D, although the side chain carboxyl oxygens of 30D have no hydrogen bonds with NFV. D30N cancels those hydrogen bonds and decreases the electrostatic interaction energy greatly. In addition, we find that the D30N mutation loses its ability to keep any water molecule at the S2 pocket. Then, the substitution of asparagine (N) for aspartate (D) at codon 30 is concluded to

TABLE 3: Binding Free Energies of the Wild-type and Mutants

	ΔG_{int}^{ele} (kcal/mol)	ΔG_{int}^{vdw} (kcal/mol)	ΔG_{sol}^{nonpol} (kcal/mol)	ΔG_{sol}^{ele} (kcal/mol)	$\Delta G_{int+sol}^{ele}$ (kcal/mol)	ΔG_b[a] (kcal/mol)	resistance level[b]
wild	−26.5 ± 4.2	−66.4 ± 3.3	−4.5 ± 0.1	44.2 ± 3.4	17.7 ± 3.5	−53.2 ± 4.2	
D30N	−19.2 ± 2.8	−64.0 ± 2.8	−4.5 ± 0.2	38.2 ± 3.6	19.0 ± 3.7	−49.5 ± 3.5	6
N88D	−26.2 ± 3.5	−66.0 ± 3.5	−4.7 ± 0.3	42.6 ± 3.7	16.4 ± 3.6	−54.3 ± 3.9	0.6
D30N/N88D	−15.1 ± 3.9	−64.9 ± 3.3	−4.6 ± 0.2	35.4 ± 3.7	20.3 ± 3.8	−49.2 ± 4.0	6
L90M	−21.1 ± 3.5	−62.2 ± 3.0	−4.2 ± 0.2	36.4 ± 3.5	15.3 ± 4.0	−51.2 ± 4.0	5

[a] $T\Delta S$ is not included. [b] References 8 and 9.

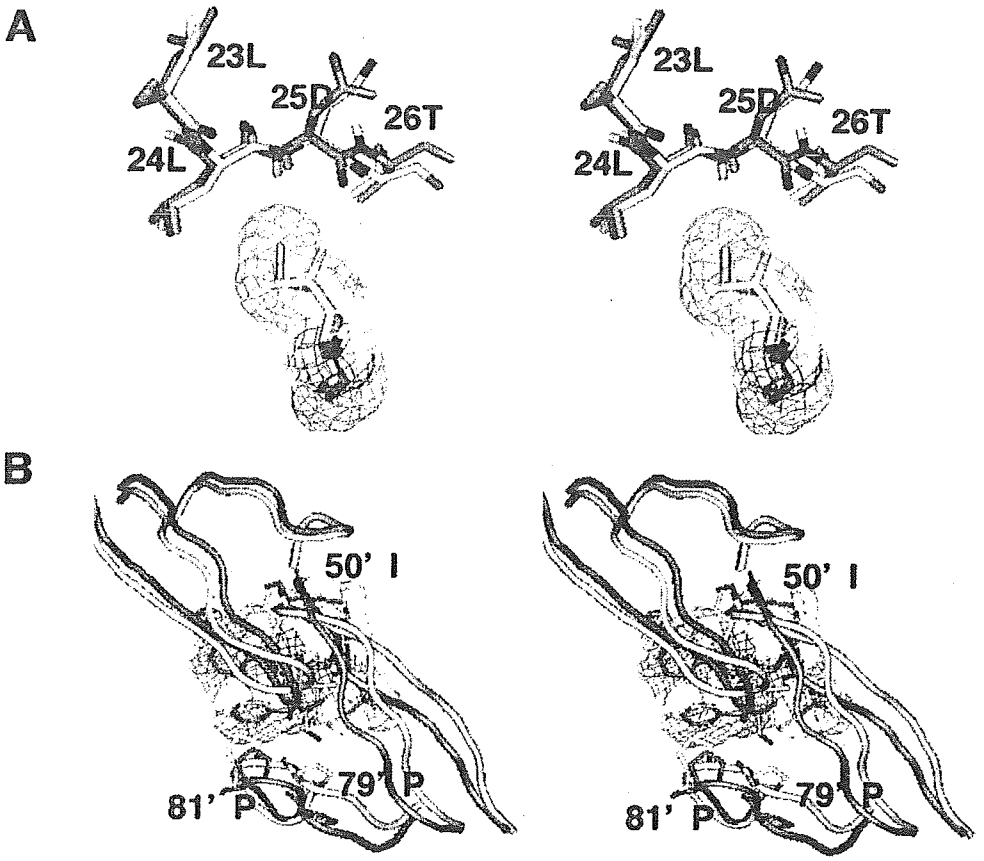


Figure 7. Stereoview of the structures (A) at the surrounding region of the 90th residue and (B) at the flap region. The wild-type structure is represented by white tubes, and L90M mutant is represented by blue tubes.

cause NFV resistance as a result of the serious decrease in the electrostatic interaction.

N88D is known as a secondary mutation for NFV and is frequently seen in the clinical scene next to D30N. Because the 88th residue is located at the helix region near the dimer interface, not at the active site, it is not clear why N88D substitution affects the resistance against NFV. Mahalingam and co-workers determined the X-ray crystal structure of the N88D mutant with substrate analogue inhibitors, not NFV.[14-16] They found that 88N in the wild-type had the proton and/or water-mediated interactions with 29D, 31T, and 74T in each subunit, whereas the corresponding water molecules were missing in both subunits of the N88D single mutant. It is also found from our calculation on the wild-type protease that the side chain of 88N makes a hydrogen bond network with 74T via one water molecule, the backbone nitrogen of 88N has a hydrogen bond with 29D, and the side chain oxygen of 88N interacts with 31T. These hydrogen bond networks are also observed in the opposite subunit. In contrast, in each of the N88D and D30N/N88D mutants, 88D interacts with only 31T and 29D, and the hydrogen bond mediating water molecules disappear in the respective subunit. That is, the 88th residue cannot form any hydrogen bonds with the 74th residue. This disappearance of the water-mediated hydrogen bond allows a large conformational change at the 74th residue. The conformational change at the β sheet consisting of the 74th residue and its adjacent contiguous residues induces the conformational changes at the neighboring β sheet and at the outside loop neighbor to the 74th residue. Further, we have detected that the slight conformational change of the NFV binding pocket due to N88D mutation induces the increase of hydrophobic contacts between NFV and the protease. The loss of the interaction between 74T and 88D would lower the constraint at 29D, 31T, and 74T, which results in the slight conformational change of the active site. Energetic analysis evidently indicates this increase of hydrophobicity. Consequently, NFV is stabilized by the increase of hydrophobic effects when N88D is acquired, while D30N/N88D destabilizes NFV by a large loss of electrostatic interaction. The reduction of the constraint might induce the compensation for the loss of replicative capacity resulting from D30N and keep the binding affinity with substrates, while the resistance against NFV is retained.

90L/M is also located at the helix region near the dimer interface, not at the active site. L90M mutants complexed with a substrate analogue inhibitor, not NFV, were also investigated by Mahalingam and co-workers.[14-16] They concluded that L90M altered van der Waals interactions in the hydrophobic interior at the dimer interface near the catalytic aspartates, and 90L related with the stability of the dimer. We also find the alteration of van der Waals interactions in the L90M model. Side chains of the 90th residues are close to the side chains of 24L/24'L and the main chain of 25D/25'D (Figure 7A). A methionine has a long and straight side chain, while a leucine has a diverging side chain. The substitution of methionine for leucine makes a collision of the 90th residues with 24L/24'L and 25D/25'D. This causes the conformational change at the triads. The conformational change of the triads induces large conformational changes at the flap and at the loop near the 79'P-81'P as a result of the presence of NFV (Figure 7B). Those regions are surprisingly very far apart from the 90th residue. In the L90M

mutant, the conformation in the binding pocket is greatly changed. Specifically, the loop region moves outward from NFV and makes a large gap; therefore, the van der Waals and hydrophobic energies decrease greatly. The character of this conformational change resembles the conformational change of the indinavir-bound L90M protease[17] and the saquinavir-bound G48V/L90M protease.[57] The reason for the multidrug resistance relevant to the L90M mutation would be the conformational change of the triads, and, subsequently, these are hydrophobic at the flap and 79′P loop. The flap and loop region interact with each of the dodecahydroisoquinoline ring of NFV, the pyridyl group of indinavir, and the planar quinoline group of saquinavir. Those chemical bases are the largest in each inhibitor and are very bulky compared with the protease substrates. No structural distortion appears at the loop region of the L90M complexed with the substrate analogue inhibitors. Therefore, it is speculated that the structural distortion seen in the inhibitor-bound L90M mutants is due to the largeness of the quinoline ring or pyridyl group in volume. Then, to reduce or eliminate the resistance of L90M, the moiety that interacts with the loop region should be less bulky.

Our MD studies indicate that the drug-resistant mutations affect the conformations of the binding cavity and the hydrophilic and hydrophobic interactions at the active site, even though the location of mutated residues is apart from the active site.

Furthermore, the difference in the binding energy between the wild-type model and those of each mutant are compatible with the indices of resistance levels that were estimated from experimental IC90.[8,9] At present, there exist some computational approaches to explain the phenotype results.[32,35,58,59] Although each of them successfully predicted the decrease of the binding affinity in the case of the active site mutation, they failed in the prediction of the nonactive site mutation as L90M. Computational prediction is usually based on the assumption that PI resistance is primarily determined by a reduction in binding affinity. Therefore, the previous studies proposed that the drug resistance due to the nonactive site mutation might be caused by another mechanism, such as decreasing the dimer stability of the protease.[57,60] However, our study indicates that the assumption is applicable to the nonactive site mutation. Some nonactive site residues without any direct contact with inhibitors (e.g., 10L, 46M, and 90L) have a strong positional correlation with the residues located at the active site. Hence, the nonactive site mutations would cause a displacement of the active site residues and the decrease of the inhibitor or substrate binding affinity. We suggest that evaluating the positions of all the residues in the mutated protease is a key factor for the success in computational prediction of the protease resistivity against PIs. Additionally, it is also applicable to the design to reduce or to eliminate the resistance at the nonactive site mutations.

Conclusions

We executed MD simulations for the wild-type and D30N, N88D, D30N/N88D, and L90M mutants to clarify the resistant mechanism of each mutation against NFV. Our results could reproduce the phenotype data and clarified the conformational alterations at the active site and the interaction changes due to the mutation. D30N induces the disappearance of the hydrogen bond between the *m*-phenol group of NFV and the 30th residue, which results in the decrease of the electrostatic binding energy. Further, D30N loses the ability to retain a water molecule at the S2 pocket. N88D alters the conformation at the β sheets consisting of 74T and its vicinity greatly. N88D also affects the active site conformation, which creates more favorable hydrophobic binding cavity. L90M affects the triads 25D26T27G and causes subsequent large conformational changes at the flap region and the 79′P loop, though the 90th residue is far apart from those regions. L90M decreases the van der Waals binding energy greatly.

Acknowledgment. This work was supported by the Health and Labor Science Research Grant for Research on HIV/AIDS from Ministry of Health and Labor of Japan.

Supporting Information Available: RMSD of the main chain atoms compared with the X-ray crystal structure (Figure 1S). This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Kohl, N.E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A.; Scolnick, E. M.; Sigal, I. S. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 4686.

(2) Debouck, C. *AIDS Res. Hum. Retroviruses* **1992**, *8*, 153.

(3) Roberts, N. A.; Martin, J. A.; Kinchington, D.; Broadhurst, A. V.; Craig, J. C.; Duncan, I. B.; Galpin, S. A.; Handa, B. K.; Kay, J.; Krohn, A.; Lambert, R. W.; Merrett, J. H.; Mills, J. S.; Parkes, K. E. B.; Redshaw, S.; Ritchie, A. J.; Taylor, D. L.; Thomas, G. J.; Machin, P. J. *Science* **1990**, *248*, 358.

(4) Johnson, V. A.; Brun-Vézinet, F.; Clotet, B.; Conway, B.; D'Aquila, R. T.; Demeter, L. M.; Kuritzkes, D. R.; Pillay, D.; Schapiro, J. M.; Telenti, A.; Richman, D. *Top. HIV Med.* **2003**, *11*, 215.

(5) Wu, T. D.; Schiffer, C. A.; Gonzales, M. J.; Taylor, J.; Kantor, R.; Chou, S.; Israelski, D.; Zolopa, A. R.; Fessel, W. J.; Shafer, R. W. *J. Virol.* **2003**, *77*, 4836.

(6) Kantor, R.; Fessel, W. J.; Zolopa, A. R.; Israelski, D.; Shulman, N.; Montoya, J. G.; Harbour, M.; Schapiro, J. M.; Shafer, R. W. *Antimicrob. Agents Chemother.* **2002**, *46*, 1086.

(7) Condra, J. H.; Schleif, W. A.; Blahy, O. M.; Gabryelski, L. J.; Graham, D. J.; Quintero, J. C.; Rhodes, A.; Robbins, H. L.; Roth, E.; Shivaprakash, M.; Titus, D.; Yang, T.; Teppler, H.; Squires, K. E.; Deutsch, P. J.; Emini, E. A. *Nature* **1995**, *374*, 569.

(8) Patick, A. K.; Duran, M.; Cao, Y.; Shugarts, D.; Keller, M. R.; Mazabel, E.; Knowles, M.; Chapman, S.; Kuritzkes, D. R.; Markowitz, M. *Antimicrob. Agents Chemother.* **1998**, *42*, 2637.

(9) Patick, A. K.; Mo, H.; Markowitz, M.; Appelt, K.; Wu, B.; Musick, L.; Kalish, V.; Kaldor, S.; Reich, S.; Ho, D.; Webber, S. *Antimicrob. Agents Chemother.* **1996**, *40*, 292 (Erratum, *40*, 1575).

(10) Jacobson, H.; Hängii, M.; Ott, M.; Duncan, I. B.; Owen, S.; Andreoni, M.; Vella, S.; Mous, J. *J. Infect. Dis.* **1996**, *173*, 1379.

(11) Sugiura, W.; Matsuda, Z.; Yokomaku, Y.; Hertogs, K.; Larder, B.; Oishi, T.; Okano, A.; Shiino, T.; Tatsumi, M.; Matsuda, M.; Abumi, H.; Takata, N.; Shirahata, S.; Yamada, K.; Yoshikura, H.; Nagai, Y. *Antimicrob. Agents Chemother.* **2002**, *46*, 708.

(12) Sugiura, W.; Oishi, T.; Okano, A.; Matsuda, M.; Abumi, H.; Yamada, K.; Koike, M.; Taki, M.; Ishikawa, M.; Miura, T.; Hukutake, K.; Gouchi, K.; Ajisawa, A.; Iwamoto, A.; Hanabusa, H.; Mimaya, J.; Takamatsu, J.; Takata, N.; Kakishita, E.; Higasa, S.; Kashiwagi, S.; Shirahata, A.; Nagai, Y. *Jpn. J. Infect. Dis.* **1999**, *52*, 175.

(13) Kaldor, S. W.; Kalish, V. J.; Davies, J. F.; Shetty, B. V.; Fritz, J. E.; Appelt, K.; Burgess, J. A.; Campanale, K. M.; Chirgadze, N. Y.; Clawson, D. K.; Dressman, B. A.; Hatch, S. D.; Khalil, D. A.; Kosa, M. B.; Lubbehusen, P. P.; Muesing, M. A.; Patick, A. K.; Reich, S. H.; Su, K. S.; Tatlock, J. H. *J. Med. Chem.* **1997**, *40*, 3979.

(14) Mahalingam, B.; Boross, P.; Wang, Y.-F.; Louis, J. M.; Fischer, C. C.; Tozser, J.; Harrison, R. W.; Weber, I. T. *Proteins* **2002**, *48*, 107.

(15) Mahalingam, B.; Louis, J. M.; Hung, J.; Harrison, R. W.; Weber, I. T. *Proteins* **2001**, *43*, 455.

(16) Mahalingam, B.; Louis, J. M.; Reed, C. C.; Adomat, J. M.; Krouse, J.; Wang, Y.-F.; Harrison, R. W.; Weber, I. T. *Eur. J. Biochem.* **1999**, *263*, 238.

(17) Mahalingam, B.; Wang, Y.-F.; Boross, P. I.; Tozser, J.; Louis, J. M.; Harrison, R. W.; Weber, I. T. *Eur. J. Biochem.* **2004**, *271*, 1516.

(18) Piana, S.; Bucher, D.; Carloni, P.; Rothlisberger, U. *J. Phys. Chem. B* **2004**, *108*, 11139.

(19) Piana, S.; Parrinello, M.; Carloni, P. *J. Mol. Biol.* **2002**, *319*, 567.

(20) Trylska, J.; Bala, P.; Geller, M.; Grochowski, P. *Biophys. J.* **2002**, *83*, 794.

(21) Okimoto, N.; Kitayama, K.; Hata, M.; Hoshino, T.; Tsuda, M. *THEOCHEM* **2001**, *543*, 53.
(22) Okimoto, N.; Tsukui, T.; Kitayama, K.; Hata, M.; Hoshino, T.; Tsuda, M. *J. Am. Chem. Soc.* **2000**, *122*, 5613.
(23) Park, H.; Suh, J.; Lee, S. *J. Am. Chem. Soc.* **2000**, *122*, 3901.
(24) Okimoto, N.; Tsukui, T.; Hata, M.; Hoshino, T.; Tsuda, M. *J. Am. Chem. Soc.* **1999**, *121*, 7349.
(25) Ventrini, A.; López-Ortiz, F.; Alvarez, J. M.; Gonzalez, J. *J. Am. Chem. Soc.* **1998**, *120*, 1110.
(26) Liu, H.; Müller-Plathe, F.; Van Gusteren, W. F. *J. Mol. Biol.* **1996**, *118*, 3946.
(27) Weber, I. T.; Harrison, R. W. *Protein Eng.* **1996**, *9*, 679.
(28) Harrison, R. W.; Weber, I. T. *Protein Eng.* **1994**, *7*, 1353.
(29) Beveridge, A. J.; Heywood, G. C. *Biochemistry* **1993**, *32*, 3325.
(30) Goldblum, A. *Biochemistry* **1988**, *27*, 1653.
(31) Perryman, A. L.; Lin, J.-H.; McCammon, J. A. *Protein Sci.* **2003**, *13*, 1108.
(32) Shenderovich, M. D.; Kagan, R. M.; Heseltine, P. N. R.; Ramnarayan, K. *Protein Sci.* **2003**, *12*, 1706.
(33) Clemente, J. C.; Hemmrajani, R.; Blum, L. E.; Goodenow, M. M.; Dunn, B. M. *Biochemistry* **2003**, *42*, 15029.
(34) Piana, S.; Carloni, P.; Rothlisberger, U. *Protein Sci.* **2002**, *11*, 2393.
(35) Rick, S. W.; Topol, I. A.; Erickson, J. W.; Burt, S. K. *Protein Sci.* **1998**, *8*, 1750.
(36) Harte, W. E., Jr.; Beveridge, D. L. *J. Am. Chem. Soc.* **1993**, *115*, 1231.
(37) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*; Gaussian, Inc.: Pittsburgh, PA, 1998.
(38) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357.
(39) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 7*; University of California: San Francisco, 2002.
(40) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
(41) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(42) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(43) Zoete, V.; Michielin, O.; Karplus, M. *J. Mol. Biol.* **2002**, *315*, 21.
(44) Schneider, T. R. *Acta Crystallogr., Sect. D* **2002**, *58*, 195.
(45) Schneider, T. R. *Acta Crystallogr., Sect. D* **2000**, *56*, 714.
(46) Williams, T.; Kelley, C. *GNUPLOT*, 1998 (contact for further information http://www.gnuplot.info/).
(47) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.
(48) Connolly, M. L. *J. Appl. Crystallogr.* **1983**, *16*, 548.
(49) Prabu-Jeyabalen, M.; Nalivaika, E. A.; King, N. M.; Schiffer, C. A. *J. Virol.* **2003**, *77*, 1306.
(50) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395.
(51) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144.
(52) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978.
(53) Karplus, M.; Petsko, G. A. *Nature* **1990**, *347*, 631.
(54) Kuhn, L. A.; Siani, M. A.; Pique, M. E.; Fisher, C. L.; Getzoff, E. D.; Tainer, J. A. *J. Mol. Biol.* **1992**, *228*, 13.
(55) Chothia, C. *J. Mol. Biol.* **1976**, *105*, 1.
(56) Nozaki, Y.; Tanford, C. *J. Biol. Chem.* **1971**, *7*, 2211.
(57) Hong, L.; Zhang, X. C.; Hartsuck, J. A.; Tang, J. *Protein Sci.* **2000**, *9*, 1898.
(58) Jenwitheesuk, E.; Samudrala, R. *BMC Struct. Biol.* **2002**, *3*, 2.
(59) Wang, W.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14937.
(60) Xie, D.; Gulnik, S.; Gustchina, E.; Yu, B.; Shao, W.; Qoronfleh, W.; Nathan, A.; Erickson, J. W. *Protein Sci.* **1999**, *8*, 1702.

ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, Feb. 2006, p. 694–701 Vol. 50, No. 2
0066-4804/06/$08.00+0 doi:10.1128/AAC.50.2.694–701.2006

Discordances between Interpretation Algorithms for Genotypic Resistance to Protease and Reverse Transcriptase Inhibitors of Human Immunodeficiency Virus Are Subtype Dependent

Joke Snoeck,[1] Rami Kantor,[2] Robert W. Shafer,[2] Kristel Van Laethem,[1] Koen Deforche,[1] Ana Patricia Carvalho,[3] Brian Wynhoven,[4] Marcelo A. Soares,[5] Patricia Cane,[6] John Clarke,[7] Candice Pillay,[8] Sunee Sirivichayakul,[9] Koya Ariyoshi,[10] Africa Holguin,[11] Hagit Rudich,[12] Rosangela Rodrigues,[13] Maria Belen Bouzas,[14] Françoise Brun-Vézinet,[15] Caroline Reid,[16] Pedro Cahn,[14] Luis Fernando Brigido,[13] Zehava Grossman,[12] Vincent Soriano,[11] Wataru Sugiura,[10] Praphan Phanuphak,[9] Lynn Morris,[8] Jonathan Weber,[7] Deenan Pillay,[17] Amilcar Tanuri,[5] Richard P. Harrigan,[4] Ricardo Camacho,[3] Jonathan M. Schapiro,[18] David Katzenstein,[2] and Anne-Mieke Vandamme[1]*

Rega Institute for Medical Research, Leuven, Belgium[1]; Stanford University, Stanford, California[2]; Hospital Egas Moniz, Lisbon, Portugal[3]; BC Center for Excellence in HIV/AIDS, Vancouver, Canada[4]; Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil[5]; Health Protection Agency, Salisbury, United Kingdom[6]; Wright Fleming Institute, London, United Kingdom[7]; National Institute for Communicable Diseases, Johannesburg, South Africa[8]; Chulalongkorn University, Bangkok, Thailand[9]; National Institute of Infectious Diseases, Tokyo, Japan[10]; Hospital Carlos III, Madrid, Spain[11]; Ministry of Health, Tel Aviv, Israel[12]; Instituto Adolfo Lutz, Sao Paulo, Brazil[13]; Fundación Huesped, Buenos Aires, Argentina[14]; Laboratory of Virology, Bichat, Claude Bernard Hospital, Paris, France[15]; Bayer Health Care-Diagnostics, Toronto, Canada[16]; HPA Antiviral Susceptibility Reference Unit, Birmingham, United Kingdom[17]; and National Hemophilia Center, Sheba Medical Center, Tel Aviv, Israel[18]

Received 5 May 2005/Returned for modification 25 August 2005/Accepted 26 November 2005

The major limitation of drug resistance genotyping for human immunodeficiency virus remains the interpretation of the results. We evaluated the concordance in predicting therapy response between four different interpretation algorithms (Rega 6.3, HIVDB-08/04, ANRS [07/04], and VGI 8.0). Sequences were gathered through a worldwide effort to establish a database of non-B subtype sequences, and demographic and clinical information about the patients was gathered. The most concordant results were found for nonnucleoside reverse transcriptase (RT) inhibitors (93%), followed by protease inhibitors (84%) and nucleoside RT inhibitor (NRTIs) (76%). For therapy-naive patients, for nelfinavir, especially for subtypes C and G, the discordances were driven mainly by the protease (PRO) mutational pattern 82I/V + 63P + 36I/V for subtype C and 82I + 63P + 36I + 20I for subtype G. Subtype F displayed more discordances for ritonavir in untreated patients due to the combined presence of PRO 20R and 10I/V. In therapy-experienced patients, subtype G displayed a lot of discordances for saquinavir and indinavir due to mutational patterns involving PRO 90 M and 82I. Subtype F had more discordance for nelfinavir attributable to the presence of PRO 88S and 82A + 54V. For the NRTIs lamivudine and emtricitabine, CRF01_AE had more discordances than subtype B due to the presence of RT mutational patterns 65R + 115 M and 118I + 215Y, respectively. Overall, the different algorithms agreed well on the level of resistance scored, but some of the discordances could be attributed to specific (subtype-dependent) combinations of mutations. It is not yet known whether therapy response is subtype dependent, but the advice given to clinicians based on a genotypic interpretation algorithm differs according to the subtype.

Genotyping for the assessment of anti-human immunodeficiency virus (HIV) drug resistance is often used in the management of individual patient therapy. Currently, it is recommended in European as well as American guidelines (17, 38). In several retrospective and prospective studies, genotyping proved beneficial in optimizing treatment for individual patients (5, 10, 16, 23, 25, 31, 37).

Although genotyping is commonly used, there are still many uncertainties with respect to the value of genotype in the assignment of a new regimen. The current genotypic assays are not always able to report all drug resistance mutations among non-B subtypes (11, 18, 19, 24). Regardless of subtype, genotyping is not sensitive to mutations that are present as a minor variant in the population (22, 40). Genotyping results also differ depending on the laboratory where they are performed. Quality control studies indicate that mutations, even present as a pure variant, are often underestimated (32).

However, separate from the quality and sensitivity issues, the interpretation of genotypic results is still not standardized.

* Corresponding author. Mailing address: Rega Institute for Medical Research, Minderbroedersstraat 10, 3000 Leuven, Belgium. Phone: 32 16332160. Fax: 32 16332131. E-mail: annemie.vandamme@uz.kuleuven.ac.be.

Several interpretation algorithms have been designed to aid in this, but they may differ in the prediction of therapy response and/or drug susceptibility. Studies were performed mainly on subtype B viruses, and even within this subtype, differences have been detected (6, 21, 29, 34, 35, 36).

Non-B subtypes are a challenge for these systems, since algorithms for these subtypes were designed using genotype, phenotype, and therapy response information that was largely derived from experience with subtype B. Recent analyses suggest that non-B viruses can develop specific mutations that differ from those identified in subtype B under the same treatment pressure (1, 20). For example, in CRF01_AE but not in subtype B viruses, V75M seems to be significantly associated with stavudine treatment (2) and, in subtype C but not in subtype B, V106M is a signature substitution of patients treated with efavirenz (4). There is a continuing controversy about the impact of secondary protease mutations (positions 36, 71, 77, etc.) which evolve in subtype B following protease exposure and are relatively frequent in untreated patients with non-B subtypes. It has been suggested that some of these can affect the susceptibility to certain protease inhibitor (PI) therapies in B and non-B subtypes (14, 28).

Although some short-term studies suggest little difference in therapy response in patients carrying non-B subtypes from that of patients infected with subtype B (12), other studies showed a significant difference in responses to treatment for different subtypes (8, 13). However, current studies have included a limited number of subjects. Potential differences can be due to differences in drug resistance. It is therefore important to know how the current drug resistance interpretation systems perform on different subtypes, and first of all, we need to know what the subtype-dependant discrepancies between the systems are.

Comparisons between these interpretation systems have already been made for subtype B strains; however, the subtype dependency of resistance assessment by these interpretations systems has not yet been determined (6, 21, 29, 34, 35, 36). In this study, we investigated four frequently used interpretation systems across a large number of non-B sequences to determine whether discordance between the systems was dependent on the viral subtype.

MATERIALS AND METHODS

Sequences. Sequences of HIV-1 protease (positions 1 to 99) and reverse transcriptase (RT) (positions 1 to 240) were collected from the published literature and from 14 laboratories in 12 countries through the non-B workgroup, a worldwide effort to establish a database of non-subtype B sequences (20). Three separate analyses were performed based on the treatment history of the patient at the time of sequencing: PI analysis, nucleoside RT inhibitor (NRTI) analysis, and nonnucleoside RT inhibitor (NNRTI) analysis. A sequence was included in the respective analysis either if the patient was reported to have had no previous exposure to a drug in that class or if the patient was being treated with a drug in that class at the time of sequencing, thus separating the analyses according to drug class exposure. In this way, sequences from patients that had drug exposure from a particular class in the past but were not at the time of sequencing taking a drug from that class were excluded. The treatment data gathered for this database were therapy history, with start and stop dates for a treatment, the regimens in the therapy, and the doses of the separate antivirals. Sequences were excluded when there was no therapy history.

Subtyping. Subtyping was performed by phylogenetic analysis using the subtyping tool developed by de Oliveira et al. separately for protease and reverse transcriptase sequences (7). Briefly, sequences are first analyzed using pure subtypes as a reference; in a second step, known circulating recombinant forms are added to the alignment. To detect recombination, bootscanning was performed using a sliding window of 400 nucleotides that was advanced 20 nucleotides at a time. Recombinants were included only if they were CRF01_AE or CRF02_AG since we had sufficient data for only these two circulating recombinant forms.

Algorithms. Four publicly available algorithms were applied on each of the sequences: Agence Nationale de Recherche sur le SIDA (ANRS) July 2004 (http://www.sante.gouv.fr/htm/actu/36_vih_2.htm) (25), HIV RT and Protease Sequence Database (HIVDB) August 2004 (http://hivdb.stanford.edu) (33), Rega Institute (Rega) version 6.3 (http://www.kuleuven.be/rega/cev/pdf/ResistanceAlgorithm6_3.pdf) (39), and Bayer Health Care-Diagnostics (VGI) version 8 (30) (formerly Visible Genetics).

Mutations considered. In all statistical analyses (see below), we scored all mutations that are included in one of the algorithms we used in the analyses: 18 NRTI resistance positions, i.e., 41, 44, 62, 65, 67, 69, 70, 74, 75, 77, 115, 116, 118, 151, 184, 210, 215, and 219; 16 NNRTI resistance positions, i.e., 98, 100, 101, 103, 106, 108, 179, 181, 188, 190, 225, 227, 230, 234, 236, and 318; and 23 PI resistance positions, i.e., 10, 20, 24, 30, 32, 33, 36, 46, 47, 48, 50, 53, 54, 60, 63, 71, 73, 77, 82, 84, 88, 90, and 93. For most positions, more than one mutant amino acid can be scored. All mixtures at resistance positions were scored as mutants.

Scoring of discordances—statistical analyses and data mining. The algorithm specification interface at the web site for the Stanford HIV drug resistance database (http://hivdb.stanford.edu) was used to apply the interpretation algorithms to each sequence (3). We assigned three levels of resistance: susceptible (S), intermediate (I), and resistant (R). For HIVDB, which assigns five levels of resistance, we obtained three by pooling the two highest and two lowest categories.

Interpretations were considered concordant if each of the algorithms assigned the same level of resistance to a sequence for a particular drug. We considered the algorithms to be fully discordant if one of them scored the sequence S for a particular drug, and another one scored it as R. Interpretations were considered partially discordant when, among the scores of the different systems, both S and I or both R and I were found for the same drug. The numbers of fully discordant (counted as 1) and partially discordant (counted as 0.5) strains were added to compute the proportion of discordant strains.

Statistical analyses were performed to see whether the number of discordances were drug and subtype dependent. We performed a one-way analysis of variance (ANOVA) with Tukey's confidence intervals to check for differences between different drugs and different subtypes. Differences between only subtype B and each of the other subtypes have been analyzed in this study.

The data mining program Weka, version 3.4.4 (http://www.cs.waikato.ac.nz/~ml/weka/), was used to identify mutational patterns that were responsible for the observed discordances, thereby also identifying the algorithms that caused the discordances. We used this tool to build binary decision trees with which it tries to predict all observed discordances. To evaluate the predictive power of the decision trees, we performed a 10-fold cross-validation. In this method, the data set is split 10-fold and the predictive performance for every subset is evaluated for a decision tree trained on the other subsets.

We built a model for each drug in which we found a statistically significant effect of subtype on discordance. We included all subtypes in the model and tried to predict discordances (three levels, concordant, discordant, and partially discordant). For each leaf in the resulting tree that predicted discordance, we calculated the subtype distribution. Fisher exact tests were performed to analyze whether a rule in the decision tree explained significantly more discordances for a particular subtype.

RESULTS

Subtype distribution. We obtained protease and/or reverse transcriptase sequences from 5,030 patients. The subtype distribution for each analysis (PI, NRTI, or NNRTI) is shown in Table 1. In total, we obtained 6,916 (3,926 from naive and 2,990 from treated patients) sequences for PI analyses, 5,689 (2,331 naive and 3,358 treated) for NRTI analyses, and 5,557 (4,208 naive and 1,349 treated) for NNRTI analyses. Twelve protease and five RT sequences were filtered out due to suspected recombination or were untypable. The majority of the sequences were of a non-B subtype except for the PI-treated and NRTI-treated class, where the prevalence of subtype B was 82% and 66%, respectively. Subtypes H, J, and K were excluded because of a limited number of sequences.

TABLE 1. Subtype distribution for sequences in the analysis groups PI, NRTI, and NNRTI

Subtype	No. of sequences for[a]:					
	PI		NRTI		NNRTI	
	Naive	Treated	Naive	Treated	Naive	Treated
A	363	35	318	105	217	206
B	1,661	2,467	632	2,224	2,139	585
C	672	201	644	339	805	178
D	260	37	201	89	159	131
F	126	80	79	107	140	46
G	128	87	63	158	144	77
CRF01_AE	207	36	132	251	291	92
CRF02_AG	509	47	262	85	313	34

[a] A sequence was included in the analysis if there was no previous exposure to a drug in that class or the patient was being treated with a drug in that class at the time of sequencing.

Discordances. Overall, the different interpretation systems agreed well on the level of resistance. Eighty-four percent of the sequences had concordant results for PIs. In only 6% of the cases, the algorithms gave full discordant results; most of the observed differences were due to partial discordances (10%). For NRTIs, 76% of the sequences gave concordant results and 8% were fully discordant. The most concordant results, 93%, were found for NNRTI. Only 1% of the sequences caused full discordances. The results for each drug are shown in Fig. 1.

The concordance was significantly higher for therapy-naive patients than for treatment-experienced patients ($P < 0.0001$) for all drug classes.

Protease inhibitor analysis. The number of discordances seemed to be drug and subtype dependent for therapy-naive patients as well as treated patients (Tables 2 and 3).

In therapy-naive patients, results for nelfinavir were discordant in 1.8% of the sequences, while for lopinavir, this was 0.3% and for tipranavir, this was 0%. When considering the results for a single drug, the proportion of sequences displaying full or partial discordances was subtype dependent. Concerning specific subtypes in therapy-naive patients, discordances were observed for ritonavir (subtype F, $P < 0.01$) and nelfinavir (subtypes G and C) (Table 2).

In treated patients, the results were different. The highest level of discordance was obtained for amprenavir (50%), whereas 36% of the sequences were scored as discordant for lopinavir and 14% for nelfinavir. Tipranavir gave still the least discordant results; only 2% of the sequences were causing discordances between algorithms. Compared to subtype B, more discordances were observed for nelfinavir in subtype F and for indinavir and saquinavir in subtype G ($P < 0.01$), while less discordances were observed for amprenavir in subtypes C and D and for atazanavir in subtype C ($P < 0.01$) (Table 3).

Nonnucleoside reverse transcriptase inhibitor analysis. For therapy-naive patients, no differences could be found between drugs, while for treated patients, efavirenz scored the most discordances (11%), followed by delavirdine and nevirapine (5%).

The proportion of sequences displaying full or partial discordances was subtype dependent in this drug class except for delavirdine and nevirapine in naive patients. But no specific subtypes were found that had differences in the resistance interpretation compared to subtype B.

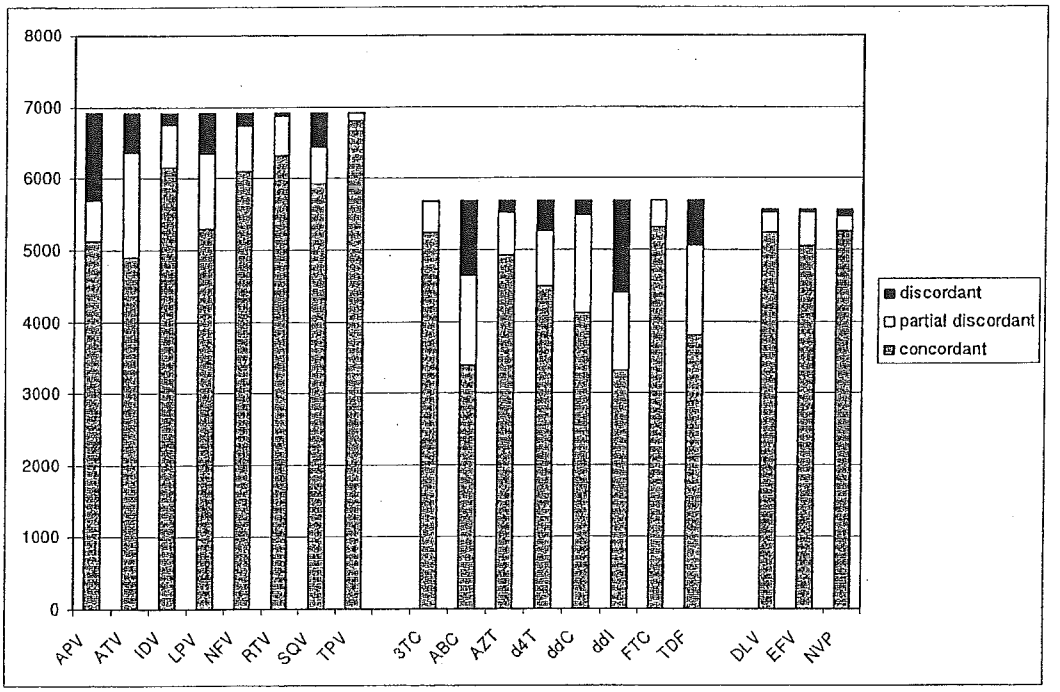


FIG. 1. Graphic representation of the number of discordant sequences per drug class. Gray bars represent the number of sequences for which concordant predictions were made by the four algorithms, white bars represent the number of sequences with partial discordance, and black bars represent sequences with discordant predictions.

TABLE 2. Interalgorithm discordances between genotypic drug resistance interpretation for sequences obtained from therapy-naive patients infected with HIV-1

Drug	Discordances (%)[a]	*P* value for subtype dependency[b]	Subtypes[c]
Protease inhibitors			
Nelfinavir	1.8	<0.01	G and C more than B
Atazanavir	1.1	<0.01	
Ritonavir	1.1	<0.01	F more than B
Amprenavir	0.6	NS	
Indinavir	0.5	NS	
Saquinavir	0.4	NS	
Lopinavir	0.3	NS	
Tipranavir	0	NS	
Nonnucleoside reverse transcriptase inhibitors			
Delavirdine	5	NS	
Nevirapine	5	NS	
Efavirenz	5	<0.01	
Nucleoside reverse transcriptase inhibitors			
Zidovudine	1.6	<0.01	
Zalcitabine	1.2	<0.01	
Stavudine	1	<0.01	C less than B
Abacavir	0.7	NS	
Didanosine	0.6	NS	
Tenofovir	0.4	NS	
Lamivudine	0.2	NS	
Emtricitabine	0.1	NS	

[a] Percentage of sequences that had discordant results between genotypic interpretation algorithms.
[b] One-way ANOVA was used to evaluate whether the number of discordances was subtype dependent (*P* of <0.05 was considered significant). NS, not significant.
[c] If the number of discordances was subtype dependent, Tukey's confidence intervals were used for a pairwise analysis to look for subtypes that caused significantly fewer or more discordances than subtype B. Although the percentage of discordances for some drugs was significantly subtype dependent, this did not always relate to a specific subtype that displayed significantly more or less discordances then subtype B.

TABLE 3. Interalgorithm discordances between genotypic drug resistance interpretation for sequences obtained from therapy-experienced patients infected with HIV-1

Drug	Discordances (%)[a]	*P* value for subtype dependency[b]	Subtypes[c]
Protease inhibitors			
Amprenavir	50	<0.01	C and D less than B
Atazanavir	42	<0.01	C less than B
Lopinavir	36	<0.01	
Saquinavir	24	<0.01	G more than B
Indinavir	15	<0.01	G more than B
Nelfinavir	14	<0.01	F more than B
Ritonavir	9	<0.01	
Tipranavir	2	NS	
Nonnucleoside reverse transcriptase inhibitors			
Efavirenz	11	<0.01	
Delavirdine	5	<0.01	
Nevirapine	5	<0.01	
Nucleoside reverse transcriptase inhibitors			
Didanosine	54	<0.01	C and D less than B
Abacavir	49	<0.01	
Tenofovir	37	<0.01	G, A, C, and D less than B
Zalcitabine	26	<0.01	C less than B
Stavudine	23	<0.01	
Zidovudine	13	<0.01	
Lamivudine	7	<0.01	CRF01_AE more than B
Emtricitabine	5	<0.01	CRF01_AE more than B

[a] Percentage of sequences that had discordant results between genotypic interpretation algorithms.
[b] One-way ANOVA was used to evaluate whether the number of discordances was subtype dependent (*P* value of <0.05 was considered significant). NS, not significant.
[c] If the number of discordances was subtype dependent, Tukey's confidence intervals were used for a pairwise analysis to look for subtypes that caused significantly fewer or more discordances than subtype B. Although the percentage of discordances for some drugs was significantly subtype dependent, this did not always relate to a specific subtype that displayed significantly more or fewer discordances then subtype B.

Nucleoside reverse transcriptase inhibitor analysis. In 1.6% of the sequences, zidovudine (AZT) was responsible for most of the discordances in therapy-naive patients; didanosine (ddI) was responsible for most of the discordances in treated patients (54%). The difference between drugs in this class was significant for both therapy-naive (Table 2) and therapy-experienced (Table 3) patients.

For zidovudine, zalcitabine, and stavudine in the naive population, the number of discordances was associated with subtype ($P < 0.01$). For only stavudine, subtype C was found to display less discordances than subtype B.

The number of discordances was significantly associated with subtype for all drugs in therapy-experienced patients ($P < 0.01$). For lamivudine and emtricitabine, CRF01_AE seemed to display significantly more discordances than subtype B. Subtypes C and D had fewer discordant interpretations for didanosine, and subtype C had also fewer for zalcitabine. For tenofovir, a lot of non-B subtypes had fewer discordant results than subtype B. This was the case for subtypes A, C, D, and G.

Mutational features of the subtype dependency. The results have been summarized in Table 4.

In therapy-naive patients among non-B subtype viruses, subtypes C and G showed partial discordances with respect to saquinavir susceptibility.

For subtype C, the most frequent pattern that caused partial discordances was a combination of protease (PRO) 82V/I + 63P + 36V/I. This pattern significantly explained more partial discordances for subtype C than for subtype B ($P < 0.0001$). This seemed due to the HIVDB interpretation algorithm. All subtype C sequences displaying this pattern also had the PRO 93L mutation. This mutation is taken into account for only nelfinavir by the HIVDB algorithm, which scores this pattern as intermediate, while all other algorithms score these sequences susceptible.

Two rules were discovered in the tree for subtype G that explained significantly more discordances than subtype B. One was a rule very similar to that for subtype C, PRO 82I + 63P + 36I ($P = 0.04$), and the other rule was PRO 82I + 63mt (any mutation) + 20I ($P = 0.01$). In practice, these rules cover the same sequences, as all subtype G sequences with the first pattern also harbor a mutation at position PRO 20 and all

TABLE 4. Mutations at least partially responsible for the subtype dependent behavior of genotypic interpretation algorithms for a drugs and algorithms responsible for the observed discordances

Drug[a]	Subtype	Mutation patterns (score)[b]	Algorithm responsible[c]
Naive population			
Nelfinavir	C	82I/V + 63P + 36I/V (SISS)	HIVDB (all sequences also 93L, taken into account by only HIVDB)
	G	82I + 63P + 36I (SISS) and 82I + 63mt + 20I (SISS)	HIVDB (high weight for 82I)
Ritonavir	F	20R + 10V/I (nrSIS)	Rega (all sequences also 36I, three secondary PI mutations scored as I by only Rega)
Treated population			
Saquinavir	G	90M + 82I (SRIR)	ANRS (does not score this as resistant)
Indinavir	G	90M + 82I + 54V (RRSI) and	HIVDB and ANRS (all sequences also 36I, pattern scored as R by HIVDB and ANRS)
		90M + 82I + 71T + 20I (RISI)	Rega (L90M not scored as R)
Nelfinavir	F	88S (RRSI) and	Rega (scores this as S)
		82A + 54V (IRRR)	ANRS (all sequences also 36I, not scored as R by ANRS)
Lamivudine	CRF01_AE	65R + 151M (IRRI)	ANRS and VGI (do not have a rule for the presence of both)
Emtricitabine	CRF01_AE	118I + 215Y (SSInr)	Rega (all sequences also 41L and 67N, 67N scored only by Rega)

[a] Only drugs for which the subtype dependence was proven and for which we found subtypes that displayed significantly more or fewer discordances than subtype B are shown. As explained in the text, the decision trees for the drugs where subtype B displayed more discordances were often too complex. Those are not included in this table.

[b] Positions at which mutations are responsible for discordances as revealed by data mining analysis. The order of the scores is shown alphabetically according to the algorithm name (ANRS, HIVDB, Rega, and VGI). Only the scoring patterns that accounted for most of the discordances (>85%) are shown. nr, no rule available for the drug.

[c] Algorithm(s) responsible for the observed discordances. Some information is provided in parentheses as to why these algorithms cause a discordance.

sequences with the second pattern also harbor a mutation at position PRO 36. Again, these discordances were due to the HIVDB algorithm, which is the only one that takes into account mutations at position PRO 20 and gives a rather high weight for the PRO 82I mutation for nelfinavir.

For ritonavir, subtype F caused more discordances than subtype B. We found a rule, PRO 20R + 10V/I, in the decision tree explaining significantly more subtype F partial discordances than those observed in subtype B. An example of the Weka decision tree with subsequent statistical analyses is shown in Fig. 2. Those subtype F sequences all had the PRO 36I mutation and thus harbored three secondary PI mutations. The Rega algorithm scores this as intermediate for ritonavir, while all other algorithms score this as susceptible.

For NRTIs, subtype B gave a lot of discordant interpretations. The rule predictive for this discordance in the decision tree was any mutation at RT 215, but this was not significant ($P = 0.07$). When examining the data, we found that the dis-

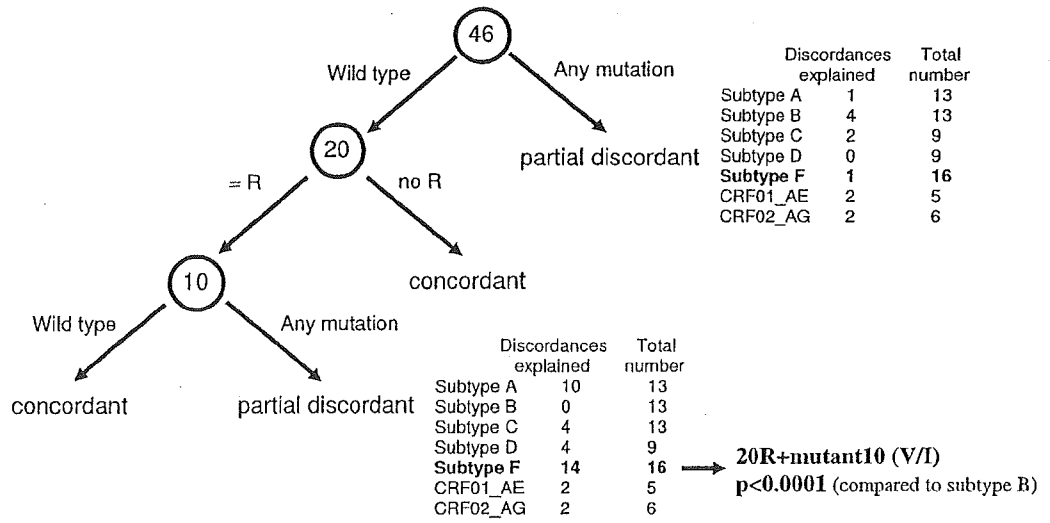


FIG. 2. Representation of the Weka decision tree for ritonavir in our untreated population. In the circles, the amino acid position is represented and, along the arrows, the mutation present is shown. R, arginine. We found that subtype F displayed more discordance. In the Weka decision tree, two rules were found, i.e., (i) any mutations at position PRO 46 and (ii) 20R + mutant 10. We calculated the number of discordances found that were explained by these rules and compared these numbers for subtype F and subtype B. Only the second rule explained significantly more discordances for subtype F than for subtype B.

cordances for stavudine were due to the ANRS system, which scores the presence of a mutation at 215 by itself as intermediately susceptible; all the other systems score this as susceptible. We found that subtype B more often had a mutation at this position than did subtype C, although this was not significant.

For the PI saquinavir in therapy-experienced patients, the full discordances observed in subtype G sequences could be attributed to mutations PRO 90 M + 82I. This was due to the ANRS interpretation system, which does not score this as resistant (as HIVDB and VGI did) if PRO 82I is present. Only PRO 82A is taken into account by ANRS.

For indinavir, subtype G also displayed more discordances than subtype B, apparently due to PRO 90 M + 82I + 54V, which was scored as resistant by HIVDB and ANRS because all these samples also had the PRO 36I mutation. Another rule predictive for discordance was PRO 90 M + 82I + 71T + 20I. The Rega system scores this pattern as susceptible, since the PRO 90 M mutation by itself is not scored as resistant by this algorithm.

Subtype F causes more discordances for nelfinavir in treated patients. The PRO 88S mutation was partially responsible for these discordances. The Rega algorithm considers these isolates to be susceptible, while the score from other algorithms was at least intermediate resistant. The partial discordances for subtype F are explained by PRO 82A + 54V. All these sequences had also PRO 36I, which is not considered resistant by ANRS relative to the other algorithms.

Subtype B displayed a lot of discordances for amprenavir. In fact, the decision tree incorporated subtype in this model. The resulting rule was PRO 90 M + 54V + 20R + 82A. All these sequences had an additional PRO 36I mutation, which is not included in the amprenavir rules of the Rega algorithm. This mutation pattern scored as intermediate for this system, while for the other algorithms, the additional PRO 36I mutation is responsible for the resistant score.

For atazanavir, subtype B caused a lot of discordances. The decision tree was very complex, and no clear rule had a high coverage and was predictive for the observed discordances in all subtypes. The atazanavir rules incorporate a number of mutations also observed for other PIs. Patients harboring a subtype B virus are probably treated with protease inhibitors more often and for a longer time, since subtype B has dominated since the beginning of the epidemic in countries where treatment was available and subsequently has been subject to drug selective pressure earlier. In these sequences, the large background of PI resistance mutations probably causes the discordances observed for atazanavir.

For lamivudine and emtricitabine (FTC), CRF01_AE scored more discordances than subtype B. For lamivudine resistance interpretation, this was caused by RT 65R + 151 M ($P < 0.05$). ANRS scores the presence of both mutations separately as intermediate but does not provide a rule for the presence of both of them, while the Rega algorithm for example scores this combination as resistant.

For emtricitabine, no clear rules were found in the tree, although it seemed that RT 41L + 67N + 118I + 215Y caused most of the partial discordances observed for CRF01_AE. The Rega algorithm is the only one that scores the RT 67N mutation for FTC. VGI does not provide rules for FTC.

For didanosine, tenofovir, and zalcitabine, subtype B had a lot more discordant interpretations than a number of non-B subtypes. The decision trees were very complex and also for these drugs, no clear rules could be deduced.

DISCUSSION

HIV genotypic information has led to an improved understanding of mutations in *pol*, which is associated with virological failure. Although resistance genotyping still has some limitations, it is often used to guide therapy start or change. One of the major problems is the interpretation of genotypic results. The knowledge on which such interpretation systems are built is based mainly on subtype B data. Considering the possible differences in therapy response in other subtypes, it would be interesting to verify whether our genotypic interpretation systems are equally valid for all subtypes. A first approach is to map discrepancies in drug resistance interpretation algorithms between subtypes and to identify which mutational patterns are responsible for such discrepancies. Such patterns can then further be investigated by, for example, in vitro mutagenesis and measuring the associated phenotype, taking into account that virus replication under drug selective pressure not only is a matter of protease and RT mutations but also is determined by the whole viral genome.

In this study, performed on sequences obtained from 5,030 patients, we investigated subtype-dependant discrepancies between four commonly used interpretation systems (Rega 6.3, HIVDB-08/04, ANRS [07/04], and VGI 8.0). The versions analyzed were the ones available to us at the time of analysis. In the meantime, updates have become available for all of these systems. None of these systems include subtype-dependant rules.

We did find drug- and subtype-dependent differences in the drug susceptibility/therapy response predictions of commonly used interpretation algorithms. We also identified mutational patterns that seemed to be partially responsible for the observed discordances.

Concordance was the lowest in the interpretation of therapy-experienced sequences, which means that it is less clear which mutations are really important for resistance development. This may explain some of the differences seen between algorithms in predicting treatment outcome (6). For lopinavir especially, the pathway towards resistance is unclear, which explains the high number of discordant results between the interpretation systems found in therapy-experienced patients (26, 27).

Our analyses revealed that the proportion of discordances between commonly used algorithms is subtype dependent for many drugs, in naive as well as in therapy-experienced patients. Concordance was higher in naive patients. However, non-B subtype sequences and subtype B sequences overall had equal numbers of resistance mutations. Both groups had mostly "wild-type" sequences. Therefore, the higher number of concordances is probably due to a larger agreement on what is a wild-type sequence.

In naive patients, discordances were found for nelfinavir (subtypes C and G). Incidentally, it is known that the pathway towards resistance for nelfinavir differs for subtypes C and G from that for subtype B. The PRO D30N mutation is not the

preferred one as in subtype B; it seems that, rather, the PRO L90M is selected (15) (P. Gomes, I. Diogo, M. F. Gonves, et al., Abstr. 9th Conf. Retrovir. Opportunistic Infect., abstr. 46, 2002). We found mutational patterns that partially explained these discordances. Those were mostly due to combinations of secondary PI mutations, which are often present as a polymorphism in non-B subtypes. Some algorithms include these mutations in their rules, while others do not. The PRO 93L mutation for example, is included by only HIVDB and not by the other systems. This mutation was present in all subtype C sequences with the pattern PRO 82I/V + 63P + 36I/V. Similarly for subtype G, the PRO 20I mutation is incorporated by only HIVDB.

For subtype F and ritonavir, the pattern PRO 20R + 10V/I also included the PRO 36I mutation. Three secondary PI mutations are scored as intermediate by only the Rega Algorithm.

For NNRTIs, we did not find any subtype-dependent discordances in resistance scoring, although some differences in resistance development have already been reported for subtype C under efavirenz treatment (2).

For NRTIs, only in naive patients did we find that the proportion of discordances is subtype dependent for stavudine. Subtype C had significantly less discordances than subtype B due to a mutation on RT 215 that occurred more frequently in subtype B sequences.

For PI resistance in treated patients, a lot of discordances are observed for subtype G in predicting resistance for saquinavir and indinavir and in subtype F for nelfinavir resistance prediction. The patterns observed here are related to a single algorithm that scores this differently. Differences often occur due to the presence of the PRO 36I mutation, which is present as a polymorphism in non-B subtypes. This mutation often triggers the switch to score an isolate as intermediate, while other systems do not take into account the substitution and consider the isolate to be susceptible. Apparently, there is no agreement on the role of some of these polymorphic resistance mutations in PI resistance.

For amprenavir and atazanavir, subtype B displayed a lot of discordances for treated patients. The decision trees for these drugs were very complex. The tree for amprenavir included subtype as a node, so a rule, PRO 90 M + 54V + 20R + 82A, could be deduced. For atazanavir, no clear rule was found. These two drugs are only recently being used in clinical practice, and the pathway towards resistance is not fully understood yet. The presence of a number of PI mutations, instead of some clear rules, is mostly used in the algorithms.

For lamivudine and emtricitabine in treated patients, CRF01_AE scored more discordances than subtype B. Although resistance for both drugs are predicted by the same rules in the algorithms, different mutation patterns are found in the decision trees. For lamivudine resistance interpretation, this was caused by RT 65R + 151 M. For emtricitabine, this was RT 41L + 67N + 118I + 215Y (although not statistically supported).

Tipranavir has a low number of discordances for naive patients as well as treated patients. This is mainly due to the limited amount of information that is available on resistance towards this drug (9). All algorithms are based on the same available information and thus predict the same level of resistance.

The four evaluated algorithms, in fact, belong to two different models. The Stanford algorithm assigns a score to each of the observed mutations and uses the sum to decide on the level of resistance, allowing complex patterns of mutations to be taken into account. The VGI, ANRS, and Rega algorithms are restrained to specific rules that describe specific mutational patterns. Therefore, the discordance for complex patterns is especially inevitable since both models use different ways to take these into account.

This study is not intended to draw conclusions on the validity of the different algorithms, but rather to identify mutation patterns that result in divergence between the algorithms, among different subtypes. The mutations and particularly the patterns of polymorphisms in non-B subtypes that are associated with viral resistance warrant further in vitro studies and ultimately need to be confirmed by clinical observation. We acknowledge, as a limitation of this study, the absence of measures of either in vitro or clinical resistance, which are phenotype and therapy outcome, respectively. However, the mutation patterns associated with discordance between the algorithms may identify the sequences of interest in larger datasets, obtained prospectively, and linked to viral load and/or CD4 data to correlate treatment outcomes.

In conclusion, the different algorithms agreed quite well on the level of resistance scored. However, where there are differences, in many cases these can be attributed to specific subtype-dependent combinations of mutations. The mutations found here should further be investigated as to whether they contribute to differences in resistance and therapy response between different subtypes. Our expertise in interpretation of genotypic resistance will increase with a scale-up of treatment to include millions of individuals with non-subtype B virus infections.

REFERENCES

1. **Abecasis, A. B., K. Deforche, J. Snoeck, L. Bacheler, P. McKenna, P. Carvalho, P. Gomes, R. Camacho, and A.-M. Vandamme.** 2005. Protease mutation M89I/V is linked to therapy failure in patients infected with the HIV-1 non-B subtypes C, F or G. AIDS **19**:1799–1806.
2. **Ariyoshi, K., M. Matsuda, H. Miura, S. Tateishi, K. Yamada, and W. Sugiura.** 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.
3. **Betts, B. J., and R. W. Shafer.** 2003. Algorithm specification interface for human immunodeficiency virus type 1 genotypic interpretation. J. Clin. Microbiol. **41**:2792–2794.
4. **Brenner, B., D. Turner, M. Oliveira, D. Moisi, M. Detorio, M. Carobene, R. G. Marlink, J. Schapiro, M. Roger, and M. A. Wainberg.** 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.
5. **Cohen, C. J., S. Hunt, M. Sension, C. Farthing, M. Conant, S. Jacobson, J. Nadler, W. Verbiest, K. Hertogs, M. Ames, A. R. Rinehart, and N. M. Graham.** 2002. A randomized trial assessing the impact of phenotypic resistance testing on antiretroviral therapy. AIDS **16**:579–588.
6. **De Luca, A., A. Cingolani, S. Di Giambenedetto, M. P. Trotta, F. Baldini, M. G. Rizzo, A. Bertoli, G. Liuzzi, P. Narciso, R. Murri, A. Ammassari, C. F. Perno, and A. Antinori.** 2003. Variable prediction of antiretroviral treatment outcome by different systems for interpreting genotypic human immunodeficiency virus type 1 drug resistance. J. Infect. Dis. **187**:1934–1943.
7. **de Oliveira, T., K. Deforche, S. Cassol, M. O. Salminen, D. Paraskevis, C. Seebregts, J. Snoeck, E. J. van Rensburg, A. M. J. Wensing, D. A. M. C. van de Vijver, C. A. Boucher, R. Camacho, and A.-M. Vandamme.** 2005. An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics **21**:3797–3800.
8. **De Wit, S., R. Boulmé, B. Poll, J.-C. Schmit, and N. Clumeck.** 2004. Viral load and CD4 cell response to protease inhibitor-containing regimens in subtype B versus non-B treatment-naive HIV-1 patients. AIDS **18**:2330–2331.

9. **Doyon, L., S. Tremblay, L. Bourgon, E. Wardrop, and M. G. Cordingley.** 2005. Selection and characterization of HIV-1 showing reduced susceptibility to the non-peptidic protease inhibitor tipranavir. Antivir. Res. **68**:27–35.
10. **Durant, J., P. Clevenbergh, P. Halfon, P. Delgiudice, S. Porsin, P. Simonet, N. Montagne, C. A. Boucher, J. M. Schapiro, and P. Dellamonica.** 1999. Drug-resistance genotyping in HIV-1 therapy: the VIRADAPT randomised controlled trial. Lancet **353**:2195–2199.
11. **Fontaine, E., C. Riva, M. Peeters, J.-C. Schmit , E. Delaporte, K. Van Laethem, K. Van Vaerenbergh, J. Snoeck, E. Van Wijngaerden, E. De Clercq, E. M. Van Ranst, and A.-M. Vandamme.** 2002. Evaluation of two commercial kits for the detection of genotypic drug-resistance on a panel of human immunodeficiency virus type-1 subtypes A-J. J. Acquir. Immune Defic. Syndr. **28**:254–258.
12. **Frater, A. J., A. Beardall, K. Ariyoshi, D. Churchill, S. Galpin, J. R. Clarke, J. N. Weber, and M. O. McClure.** 2001. Impact of baseline polymorphisms in RT and protease on outcome of highly active antiretroviral therapy in HIV-1-infected African patients. AIDS **15**:1493–1502.
13. **Frater, A. J., D. T. Dunn, A. J. Beardall, K. Ariyoshi, J. R. Clarke, M. O. McClure, and J. N. Weber.** 2002. Comparative response of African HIV-1-infected individuals to highly active antiretroviral therapy. AIDS **16**:1139–1146.
14. **Gonzalez, L. M. F., R. M. Brindeiro, M. Tarin, A. Calazans, M. A. Soares, S. Cassol, and A. Tanuri.** 2003. In vitro hypersusceptibility of human immunodeficiency virus type 1 subtype C protease to lopinavir. Antimicrob. Agents Chemother. **47**:2817–2822.
15. **Grossman, Z., E. E. Paxinos, D. Averbuch, S. Maayan, N. T. Parkin, D. Engelhard, M. Lorber, V. Istomin, Y. Shaked, E. Mendelson, D. Ram, C. J. Petropoulos, and J. M. Schapiro.** 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.
16. **Haubrich, R., and L. M. Demeter.** 2001. Clinical utility of resistance testing: retrospective and prospective data supporting use and current recommendations. J. Acquir. Immune Defic. Syndr. **26**:S51–S59.
17. **Hirsch, M. S., F. Brun-Vezinet, C. Bonaventura, B. Conwy, D. R. Kuritzkes, R. T. D'Aquila, L. M. Demeter, S. M. Hammer, V. A. Johnson, C. Loveday, J. W. Mellors, D. M. Jacobsen, and D. D. Richman.** 2003. Antiretroviral drug resistance testing in adults infected with human immunodeficiency virus type 1: 2003 recommendations of an International AIDS Society–USA Panel. Clin. Infect. Dis. **37**:113–128.
18. **Holguin, A., K. Hertogs, and V. Soriano.** 2003. Performance of drug resistance assays in testing HIV-1 non-B subtypes. Clin. Microbiol. Infect. **9**:323–326.
19. **Jagodzinski, L. L., J. D. Cooley, M. Weber, and N. L. Michael.** 2003. Performance characteristics of human immunodeficiency virus type 1 (HIV-1) genotyping systems in sequence-based analysis of subtypes other than HIV-1 subtype B. J. Clin. Microbiol. **41**:998–1003.
20. **Kantor, R., D. Katzenstein, B. Efron, P. Carvalho, B. Wynhoven, P. Cane, J. R. Clarke, S. Sirivichayakul, M. A. Soares, J. Snoeck, C. Pillay, H. Rudich, R. Rodrigues, A. Holguin, K. Ariyoshi, P. Weidle, M. B. Bouzas, P. Cahn, W. Sugiura, V. Soriano, L. F. Brigido, Z. Grossman, L. Morris, A. M. Vandamme, A. Tanuri, P. Phanuphak, J. Weber, D. Pillay, P. R. Harrigan, R. Camacho, J. M. Schapiro, and R. W. Shafer.** 26 April 2005. Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: results of a global collaboration. PLOS Med. **2**:e112. [Epub ahead of print.]
21. **Kijak, G. H., A. E. Rubio, S. E. Pampuro, C. Zala, P. Cahn, R. Galli, J. S. Montaner, and H. Salomon.** 2003. Discrepant results in the interpretation of HIV-1 drug-resistance genotypic data among widely used algorithms. HIV Med. **4**:72–78.
22. **Korn, K., H. Reil, H. Walter, and B. Schmidt.** 2003. Quality control trial for human immunodeficiency virus type 1 drug resistance testing using clinical samples reveals problems with detecting minority species and interpretation of test results. J. Clin. Microbiol. **41**:3559–3565.
23. **Loveday, C., D. Dunn, H. Green, A. R. Rinehart, and P. McKenna on behalf of the ERA Steering Committee.** 2003. A randomized controlled trial of phenotypic resistance testing in addition to genotypic resistance testing: the ERA trial. Antivir. Ther. **8**(Suppl. 1):S188.
24. **Maes, B., Y. Schrooten, J. Snoeck, I. Derdelinckx, M. Van Ranst, A. M. Vandamme, and K. Van Laethem.** 2004. Performance of Viroseq HIV-1 genotyping system in routine practice at a Belgian clinical laboratory. J. Virol. Methods **119**:45–49.
25. **Meynard, J. L., M. Vray, L. Morand-Joubert, E. Race, D. Descamps, G. Peytavin, S. Matheron, C. Lamotte, S. Guiramand, D. Costagliola, F. Brun-Vezinet, F. Clavel, P. M. Girard, and the Narval Trial Group.** 2002. Phenotypic or genotypic resistance testing for choosing antiretroviral therapy after treatment failure: a randomized trial. AIDS **16**:727–736.
26. **Monno, L., A. Saracino, L. Scudeller, G. Pastore, S. Bonora, A. Cargnel, G. Carosi, and G. Angarano.** 2003. HIV-1 phenotypic susceptibility to lopinavir (LPV) and genotypic analysis in LPV/r-naive subjects with prior protease inhibitor experience. J. Acquir. Immune Defic. Syndr. **33**:439–447.
27. **Parkin, N. T., C. Chappey, and C. J. Petropoulos.** 2003. Improving lopinavir genotype algorithm through phenotype correlations: novel mutation patterns and amprenavir cross-resistance. AIDS **17**:955–961.
28. **Perno, C. F., A. Cozzi-Lepri, F. Forbici, A. Bertoli, M. Violin, M. Stella Mura, G. Cadeo, A. Orani, A. Chirianni, C. De Stefano, C. Balotta, A. d'Arminio Monforte, and the Italian Cohort Naive Antiretrovirals Study Group.** 2004. Minor mutations in HIV protease at baseline and appearance of primary mutation 90M in patients for whom their first protease-inhibitor antiretroviral regimens failed. J. Infect. Dis. **189**:1983–1987.
29. **Ravela, J., B. J. Betts, F. Brun-Vezinet, A.-M. Vandamme, D. Descamps, K. Van Laethem, K. Smith, J. M. Schapiro, D. L. Winslow, C. Reid, and R. W. Shafer.** 2003. HIV-1 protease and reverse transcriptase mutation patterns responsible for discordances between genotypic drug resistance interpretation algorithms. J. Acquir. Immune Defic. Syndr. **33**:8–14.
30. **Reid, C. L., R. Bassett, S. Day, B. Larder, V. De Gruttola, and D. L. Winslow.** 2002. A dynamic rules-based interpretation system derived by an expert panel is predictive of virological failure. Antivir. Ther. **7**:S121.
31. **Sarmati, L., E. Nicastri, M. A. Montano, L. Dori, A. R. Buonomini, G. d'Ettorre, F. Gatti, S. G. Parisi, V. Vullo, and M. Andreoni.** 2004. Decrease of replicative capacity of HIV isolates after genotypic guided change of therapy. J. Med. Virol. **72**:511–516.
32. **Schuurman, R., D. Brambilla, T. de Groot, D. Huang, S. Land, J. Bremer, I. Benders, C. A. Boucher, and the ENVA Working Group.** 2002. Underestimation of HIV type 1 drug resistance mutations: results from the ENVA-2 genotyping proficiency program. AIDS Res. Hum. Retrovir. **18**:243–248.
33. **Shafer, R. W., R. J. Duane, B. J. Betts, Y. Xi, and M. J. Gonzales.** 2000. Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic Acids Res. **28**:346–348.
34. **Stürmer, M., H. W. Doerr, and W. Preiser.** 2003. Variety of interpretation systems for human immunodeficiency virus type 1 genotyping: confirmatory information or additional confusion? Curr. Drug Targets Infect. Disord. **3**:255–262.
35. **Stürmer, M., H. W. Doerr, S. Staszewski, and W. Preiser.** 2003. Comparison of nine resistance interpretation systems for HIV-1 genotyping. Antivir. Ther. **8**:55–60.
36. **Torti, C., E. Quiros-Roldan, W. Keulen, L. Scudeller, S. Lo Caputo, C. A. Boucher, F. Castelli, F. Mazzotta, P. Pierotti, A. M. Been-Tiktak, G. Buccolieri, M. De Gennaro, G. Carosi, C. Tinelli, and the GenPherex Study Group of the MaSTeR Cohort.** 2003. Comparison between rules-based human immunodeficiency virus type 1 genotype interpretations and real or virtual phenotype: concordance analysis and correlation with clinical outcome in heavily treated patients. J. Infect. Dis. **188**:194–201.
37. **Tural, C., L. Ruiz, C. Holtzer, J. Schapiro, P. Viciana, J. Gonzales, P. Domingo, C. A. Boucher, C. Rey-Joly, B. Clotet, and the Havana Study Group.** 2002. The clinical utility of HIV-1 genotyping and expert advice: the Havana trial. AIDS **16**:209–218.
38. **Vandamme, A. M., A. Sonnerborg, M. Ait-Khaled, J. Albert, B. Asjo, L. Bacheler, D. Banhegyi, C. A. Boucher, F. Brun-Vezinet, R. Camacho, P. Clevenbergh, N. Clumeck, N. Dedes, A. De Luca, H. W. Doerr, J. L. Faudon, G. Gatti, J. Gerstoft, W. W. Hall, A. Hatzakis, N. S. Hellmann, A. Horban, J. D. Lundgren, D. J. Kempf, D. Miller, V. Miller, T. W. Myers, C. Nielsen, M. Opravil, L. Palmisano, C. F. Perno, A. N. Phillips, D. Pillay, T. Pumarola, L. Ruiz, M. O. Salminen, J. M. Schapiro, B. Schmidt, J.-C. Schmit, R. Schuurman, E. Shulse, V. Soriano, S. Staszewski, S. Vella, R. Ziermann, and L. Perrin.** 2004. Updated European recommendations for the clinical use of HIV drug resistance testing. Antivir. Ther. **9**:829–848.
39. **Van Laethem, K., A. De Luca, A. Antinori, A. Cingolani, C. F. Perno, and A.-M. Vandamme.** 2002. A genotypic drug resistance algorithm that significantly predicts therapy response in HIV-1 infected patients. Antivir. Ther. **7**:123–129.
40. **Van Laethem, K., K. Van Vaerenbergh, J.-C. Schmit, S. Sprecher, P. Hermans, V. De Vroey, R. Schuurman, T. Harrer, M. Witvrouw, E. Van Wijngaerden, L. Stuyver, M. Van Ranst, J. Desmyter, E. De Clercq, and A.-M. Vandamme.** 1999. Phenotypic assays and sequencing are less sensitive than point mutation assays for detection of resistance in mixed HIV-1 genotypic populations. J. Acquir. Immune Defic. Syndr. **22**:107–118.

Impact of HIV-1 Subtype and Antiretroviral Therapy on Protease and Reverse Transcriptase Genotype: Results of a Global Collaboration

Rami Kantor[1*¤], David A. Katzenstein[1], Brad Efron[2], Ana Patricia Carvalho[3], Brian Wynhoven[4], Patricia Cane[5], John Clarke[6], Sunee Sirivichayakul[7], Marcelo A. Soares[8], Joke Snoeck[9], Candice Pillay[10], Hagit Rudich[11], Rosangela Rodrigues[12], Africa Holguin[13], Koya Ariyoshi[14], Maria Belen Bouzas[15], Pedro Cahn[15], Wataru Sugiura[14], Vincent Soriano[13], Luis F. Brigido[12], Zehava Grossman[11], Lynn Morris[10], Anne-Mieke Vandamme[9], Amilcar Tanuri[8], Praphan Phanuphak[7], Jonathan N. Weber[6], Deenan Pillay[16], P. Richard Harrigan[4], Ricardo Camacho[3], Jonathan M. Schapiro[1], Robert W. Shafer[1]

1 Division of Infectious Disease and Center for AIDS Research, Stanford University, Stanford, California, United States of America **2** Department of Statistics and Division of Biostatistics, Stanford University, Stanford, California, United States of America **3** Hospital Egas Moniz, Lisbon, Portugal, **4** BC Centre for Excellence in HIV/AIDS, Vancouver, British Columbia, Canada, **5** Health Protection Agency, Porton Down, United Kingdom, **6** Wright Fleming Institute, Imperial College, St. Mary's Hospital, London, United Kingdom **7** Chulalongkorn University, Bangkok, Thailand, **8** Universidade Federal do Rio de Janeiro, Brazil, **9** Rega Institute for Medical Research, Leuven, Belgium, **10** National Institute of Communicable Diseases, Johannesburg, South Africa, **11** Central Virology, Public Health Laboratories, Ministry of Health, Tel-Hashomer, Israel, **12** Instituto Adolfo Lutz, Sao Paulo, Brazil, **13** Hospital Carlos III, Madrid, Spain, **14** National Institute of Infectious Diseases, Tokyo, Japan, **15** Fundación Huesped, Buenos Aires, Argentina, **16** University College London and Health Protection Agency, London, United Kingdom

Competing Interests: The authors have declared that no competing interests exist.

Author Contributions: See Acknowledgments.

Academic Editor: David D. Ho, The Rockefeller University, United States of America

Citation: Kantor R, Katzenstein DA, Efron B, Carvalho AP, Wynhoven B, et al. (2005) Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: Results of a global collaboration. PLoS Med 2(4): e112.

Received: October 8, 2004
Accepted: March 7, 2005
Published: April 26, 2005

DOI: 10.1371/journal.pmed.0020112

Abbreviations: NRTI, nucleoside reverse transcriptase inhibitor; NNRTI, non-nucleoside reverse transcriptase inhibitor; PI, protease inhibitor; RT, reverse transcriptase; RTI, reverse transcriptase inhibitor

*To whom correspondence should be addressed. E-mail: rkantor@brown.edu

¤Current affiliation: Division of Infectious Diseases, Brown University, Providence, Rhode Island, United States of America

ABSTRACT

Background

The genetic differences among HIV-1 subtypes may be critical to clinical management and drug resistance surveillance as antiretroviral treatment is expanded to regions of the world where diverse non-subtype-B viruses predominate.

Methods and Findings

To assess the impact of HIV-1 subtype and antiretroviral treatment on the distribution of mutations in protease and reverse transcriptase, a binomial response model using subtype and treatment as explanatory variables was used to analyze a large compiled dataset of non-subtype-B HIV-1 sequences. Non-subtype-B sequences from 3,686 persons with well characterized antiretroviral treatment histories were analyzed in comparison to subtype B sequences from 4,769 persons. The non-subtype-B sequences included 461 with subtype A, 1,185 with C, 331 with D, 245 with F, 293 with G, 513 with CRF01_AE, and 618 with CRF02_AG. Each of the 55 known subtype B drug-resistance mutations occurred in at least one non-B isolate, and 44 (80%) of these mutations were significantly associated with antiretroviral treatment in at least one non-B subtype. Conversely, of 67 mutations found to be associated with antiretroviral therapy in at least one non-B subtype, 61 were also associated with antiretroviral therapy in subtype B isolates.

Conclusion

Global surveillance and genotypic assessment of drug resistance should focus primarily on the known subtype B drug-resistance mutations.

Introduction

The HIV-1 pandemic resulted from the cross-species transmission of a lentivirus, most likely of chimpanzee origin, that began spreading among humans during the first half of the previous century [1,2,3]. The progeny of this zoonotic infection—designated HIV-1 group M (main) viruses—make up the vast majority of HIV-1 infections. During their spread among humans, group M viruses have developed an extraordinary degree of genetic diversity, and most can be segregated into nine pure subtypes and several commonly circulating recombinant forms [4].

HIV-1 subtype B is the predominant subtype in North America, Western Europe, and Australia. The antiretroviral drugs used to treat HIV were developed using biophysical, biochemical, and in vitro studies of subtype B isolates, and most data on the genetic mechanisms of HIV-1 drug resistance are from subtype B viruses. However, HIV-1 subtype B viruses account for only approximately 12% of the global HIV pandemic [5], and as therapy is introduced into developing countries, the number of persons with non-B viruses initiating therapy will increase dramatically.

HIV-1 subtypes differ from one another by 10%–12% of their nucleotides and 5%–6% of their amino acids in protease and reverse transcriptase (RT) [6]. Intersubtype nucleotide differences influence the spectrum of amino acid substitutions resulting from point mutations, and intersubtype amino acid differences influence the biochemical and biophysical microenvironment within the protease and RT [7,8]. These differences among subtypes therefore could influence the spectrum of mutations that develop during selective drug pressure.

An increasing number of observational studies, in vitro and in vivo, suggest that the currently available protease and RT inhibitors are as active against non-B viruses as they are against subtype B viruses [9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26]. However, fewer data are available on the genetic mechanisms of drug resistance in non-B viruses, and some in vitro and in vivo observations suggest that the various subtypes may respond differently to certain antiretroviral drugs [27,28,29,30,31,32,33,34,35].

Identifying the relevant drug-resistance mutations among non-B subtypes will be important for monitoring the evolution and transmission of drug resistance, for determining initial treatment strategies for persons infected with non-B viruses, and for interpreting genetic resistance among patients who fail antiretroviral therapy.

In this study, we characterize protease and RT mutations in non-B HIV-1 subtypes from persons receiving antiretroviral therapy, and attempt to answer the following two questions. (i) Do the mutations that cause drug resistance in subtype B viruses also develop in non-B viruses exposed to antiretroviral drugs? (ii) Do novel mutations emerge in non-subtype-B viruses during antiretroviral drug failure that are not recognized in subtype B viruses?

Methods

HIV-1 Sequences and Antiretroviral Treatments

Sequences of HIV-1 protease (positions 1–99) and RT (positions 1–240) from persons whose antiretroviral treatment history was known were collected from the published literature and from 14 laboratories in 12 countries. Persons were considered untreated if they had never been exposed to antiretroviral drugs, and treated if they were receiving RT inhibitors (RTIs) and/or protease inhibitors (PIs) at the time the isolate was obtained. Sequences from treated persons were included for analysis only if they were obtained from persons whose entire treatment histories were known. If multiple isolates from the same person were sequenced, only the latest isolate was included for analysis. Only sequences determined using dideoxyterminator sequencing were included in the analysis. In all, 99% of sequences were determined using direct PCR (population-based sequencing), and 1% of sequences represented the consensus sequence of multiple clones.

Samples obtained from patients were submitted to clinical and research laboratories for resistance testing in the course of evaluation and care of HIV infection. Data analyzed included published and presented data obtained under protocols approved by national and local institutional review boards or ethical review panels in each country. Sequence, demographic, and treatment data, unlinked from all personal identifiers, were analyzed at Stanford University under a protocol approved by the Stanford University Panel on Human Subjects.

Subtype Assignment

Similarity plotting and bootscanning using a window size of 400 nucleotides and a step size of 40 nucleotides were performed using reference sequences for each of the nine pure subtypes (A, B, C, D, F, G, H, J, and K) and two recombinant forms (CRF01_AE and CRF02_AG) [36]. Isolates that contained a combination of more than one subtype were excluded from analysis, except when subtypes A and G were detected in a pattern consistent with CRF02_AG. Because CRF01_AE *pol* sequences do not contain recombinant breakpoints, subtype assignment was based on the fact that *pol* CRF01_AE and pure A sequences are divergent. This approach had an accuracy of 96% when applied to the protease and RT genes of 137 well characterized subtype A, CRF01_AE, and CRF02_AG isolates with known subtypes based on *pol* and *gag* and/or *env*, with most errors resulting from the misclassification of subtype A protease sequences as CRF01_AE (data not shown).

Reference sequences used were U455 (subtype A), CM240 (CRF01_AE), IbNG (CRF02_AG), HXB2 (subtype B), C2220 (subtype C), NDK (subtype D), 93BR020 (subtype F), SE6165 (subtype G), 90CR056 (subtype H), SE9173c (subtype J), 97EQTB11C (subtype K), YBF30 (Group N), and ANT70C (Group O). A total of 223 protease and 307 RT sequences of indeterminate subtype were excluded from the analysis.

Mutation Definitions

Each sequence was translated and compared to the consensus B protease and RT sequences in the Los Alamos HIV Sequence database (http://hiv-web.lanl.gov) using the HIVSeq program [37]. Mutations were defined as differences from the wild-type consensus B sequence. Known subtype B drug-resistance mutations were defined as follows: 18 nucleoside RTI (NRTI)–resistance positions at 41, 44, 62, 65, 67, 69, 70, 74, 75, 77, 115, 116, 118, 151, 184, 210, 215, and 219; 15 non-nucleoside RTI (NNRTI)–resistance positions at 98, 100,

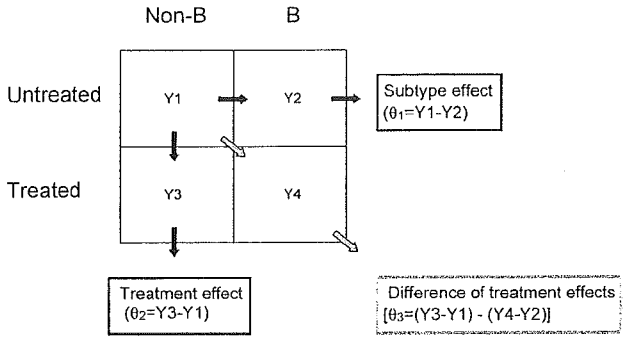


Figure 1. Binomial Response Model Used to Evaluate Subtype and Treatment Effects on Genotypic Evolution for Each Protease and RT Position

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 θ_1, the score differences between non-B and B subtypes in untreated persons. The treatment effect was evaluated by calculating θ_2, the score differences between treated and untreated persons within the same subtype. The subtype–treatment interaction was evaluated by calculating θ_3, the difference of differences in the 2 × 2 table, or the difference in treatment effects between non-B and B subtypes.

DOI: 10.1371/journal.pmed.0020112.g001

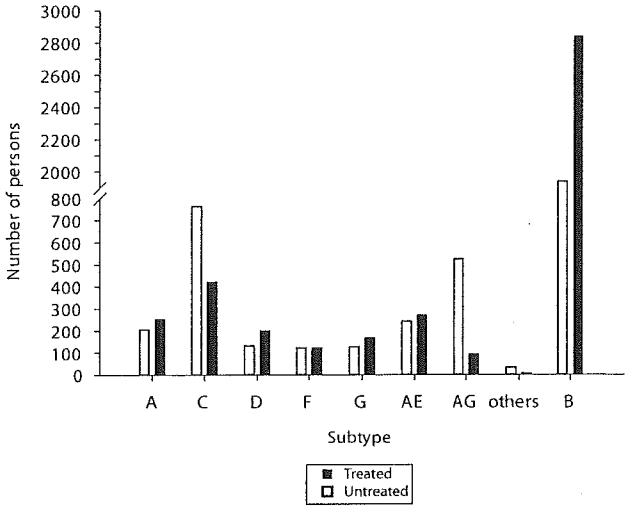


Figure 2. Number of Treated and Untreated Persons Infected with B and Non-B HIV-1 Subtypes from Whom Protease and/or RT Sequences Were Available for Analysis.

DOI: 10.1371/journal.pmed.0020112.g002

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 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[a]	Cameroon (14), Other (26)

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

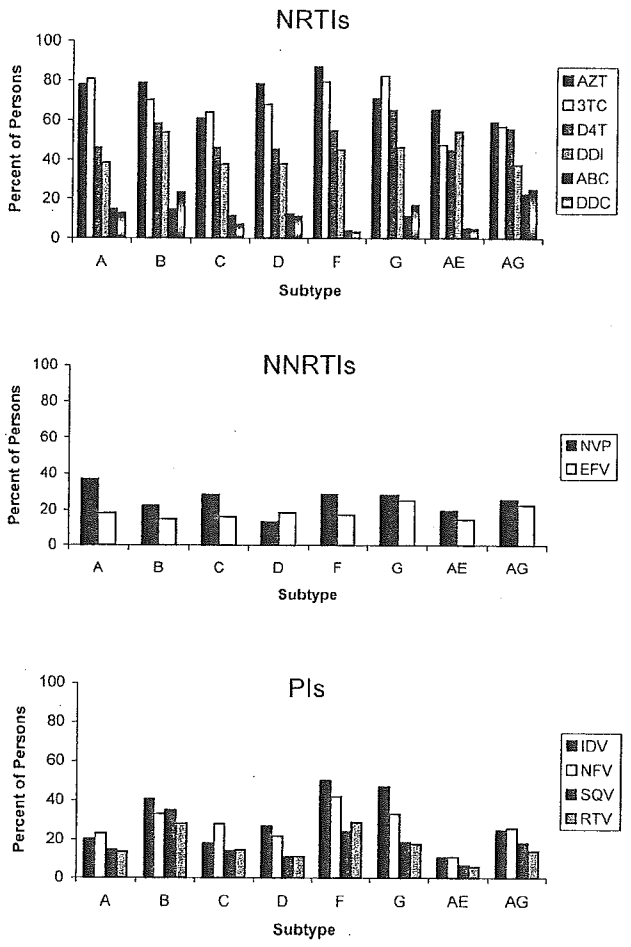


Figure 3. Proportions of Persons Receiving Treatment with Specific NRTIs, NNRTIs, and PIs

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 (θ_1), comparing codons between untreated persons infected with B and each of the non-B subtypes; (ii) a treatment parameter (θ_2), comparing codons between treated and untreated sequences of the same subtype; and (iii) an interaction parameter (θ_3), comparing the effect of treatment between subtype B and each of the non-B subtypes.

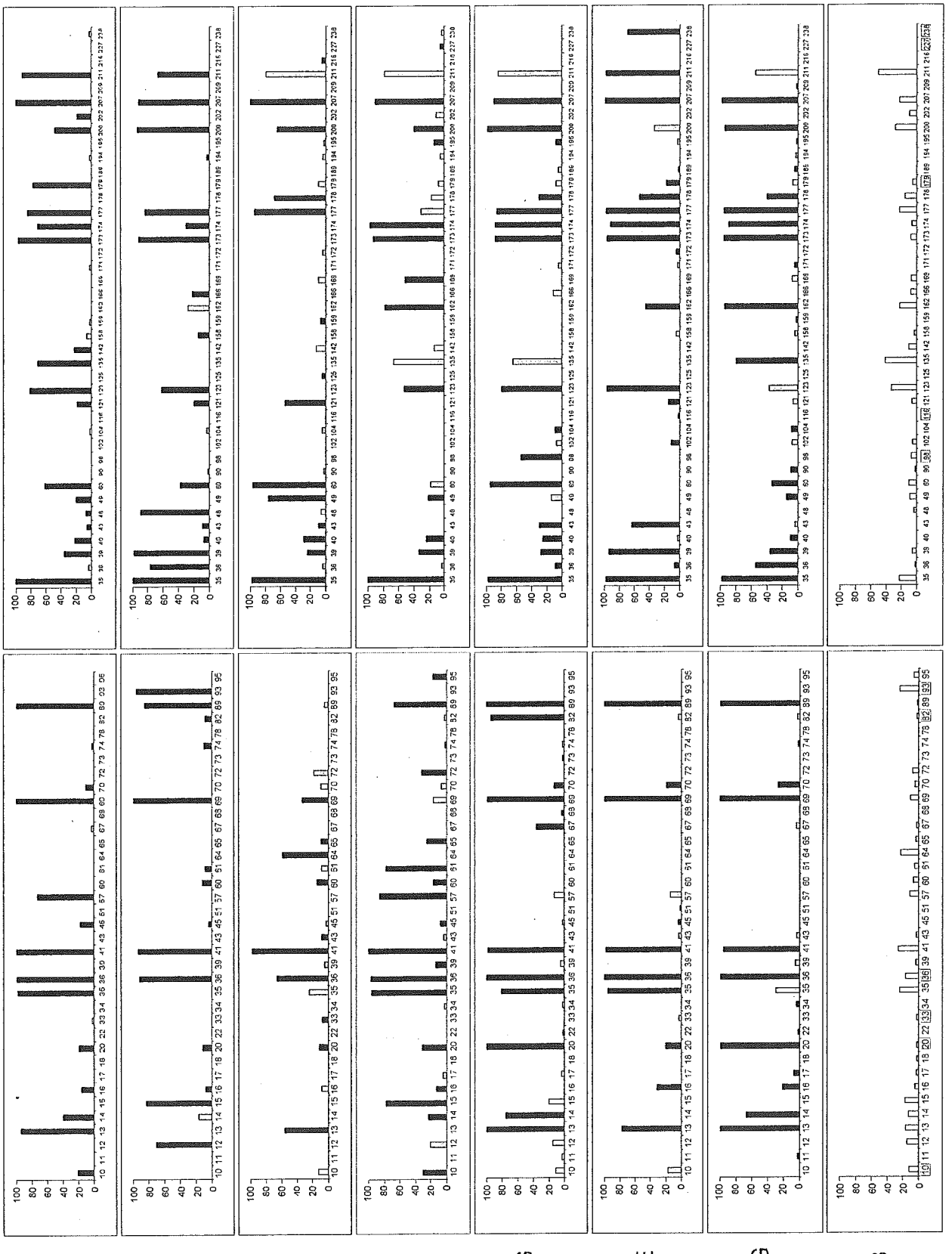
To increase the statistical power of our analysis, we made two simplifications. First, we did not distinguish between distinct substitutions at the same position; all differences from consensus B were considered mutations. Second, viruses were categorized only according to the classes of drugs (PI, NRTI,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


Protease
Polymorphisms
Reverse Transcriptase
A
C
D
F
G
AE
AG
B

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 θ_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 θ_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