

Mutation Patterns and Structural Correlates in Human Immunodeficiency Virus Type 1 Protease following Different Protease Inhibitor Treatments

Thomas D. Wu,^{1†} Celia A. Schiffer,² Matthew J. Gonzales,³ Jonathan Taylor,⁴
Rami Kantor,³ Sunwen Chou,⁵ Dennis Israelski,³ Andrew R. Zolopa,³
W. Jeffrey Fessel,⁶ and Robert W. Shafer^{3*}

Department of Biochemistry,¹ Division of Infectious Diseases, Department of Medicine,³ and Department of Statistics,⁴ Stanford University, Stanford, and AIDS Research, Kaiser-Permanente, Northern California, San Francisco,⁶ California; Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts²; and Division of Infectious Diseases, Oregon Health and Science University, Portland, Oregon⁵

Received 13 September 2002/Accepted 14 January 2003

Although many human immunodeficiency virus type 1 (HIV-1)-infected persons are treated with multiple protease inhibitors in combination or in succession, mutation patterns of protease isolates from these persons have not been characterized. We collected and analyzed 2,244 subtype B HIV-1 isolates from 1,919 persons with different protease inhibitor experiences: 1,004 isolates from untreated persons, 637 isolates from persons who received one protease inhibitor, and 603 isolates from persons receiving two or more protease inhibitors. The median number of protease mutations per isolate increased from 4 in untreated persons to 12 in persons who had received four or more protease inhibitors. Mutations at 45 of the 99 amino acid positions in the protease—including 22 not previously associated with drug resistance—were significantly associated with protease inhibitor treatment. Mutations at 17 of the remaining 99 positions were polymorphic but not associated with drug treatment. Pairs and clusters of correlated (covarying) mutations were significantly more likely to occur in treated than in untreated persons: 115 versus 23 pairs and 30 versus 2 clusters, respectively. Of the 115 statistically significant pairs of covarying residues in the treated isolates, 59 were within 8 Å of each other—many more than would be expected by chance. In summary, nearly one-half of HIV-1 protease positions are under selective drug pressure, including many residues not previously associated with drug resistance. Structural factors appear to be responsible for the high frequency of covariation among many of the protease residues. The presence of mutational clusters provides insight into the complex mutational patterns required for HIV-1 protease inhibitor resistance.

Drug resistance is a major obstacle to the effective treatment of human immunodeficiency virus type 1 (HIV-1) infection. Although 16 antiretroviral drugs have been approved for the treatment of HIV-1, cross-resistance within each of the three antiretroviral drug classes—nucleoside reverse transcriptase (RT) inhibitors, nonnucleoside RT inhibitors, and protease inhibitors—often leads to the development of multidrug resistance. HIV-1-specific protease inhibitors pose a high genetic barrier to drug resistance because multiple protease mutations are usually required for the development of resistance to these inhibitors (4, 13, 19). Nonetheless, resistance to multiple protease inhibitors occurs commonly, attesting to the conformational flexibility of the HIV-1 protease enzyme (5, 10, 13, 26).

Most of the published sequence data on protease inhibitor-associated mutations are based on isolates obtained from persons treated for no more than 1 year with a single inhibitor (4, 17, 19–21). Few published data are available from persons with carefully characterized treatment histories who have received

more than one inhibitor (12), and the genetic mechanisms by which HIV-1 protease develops resistance to multiple inhibitors have not been explored. Understanding the genetic basis of multidrug resistance, however, is critical to designing new non-cross-resistant protease inhibitors that are active against current drug-resistant HIV-1 isolates.

To characterize the patterns of mutations in protease isolates from heavily treated persons, we collected and analyzed a large number of protease sequences of HIV-1 isolates obtained from persons with a range of protease inhibitor experiences. Our analysis allows us to extend previous observations of the mutational flexibility of HIV-1 protease and to identify interactions among protease mutations. We used published structural data to explore possible underlying causes for these interactions.

MATERIALS AND METHODS

Virus isolates and sequences. We analyzed HIV-1 subtype B protease sequences from persons with well-characterized antiretroviral treatment histories. These sequences were taken from previously published studies (appearing in the 15 April 2002 release of the Stanford University HIV RT and Protease Sequence Database [<http://hivdb.stanford.edu>]) (25) and from sequencing performed at the Stanford University Hospital Diagnostic Virology Laboratory between 1 July 1997 and 31 December 2001. The isolates were subtyped by comparing them to reference sequences of known subtype (8, 15).

* Corresponding author. Mailing address: Division of Infectious Diseases, Room S-156, Stanford University, Stanford, CA 94305. Phone: (650) 725-2946. Fax: (650) 723-8596. E-mail: rshafer@stanford.edu.

† Present address: Genentech, Inc., South San Francisco, Calif.

If multiple isolates were obtained from the same person during the course of protease inhibitor treatment, we included only the most recent isolate. We included two isolates from the same person only if a pre-protease inhibitor treatment isolate was also available. Only sequences that encompassed positions 10 to 90 were included in our analysis (96% included the complete protease, positions 1 to 99). All isolates were sequenced by dideoxynucleotide sequencing rather than by hybridization assays.

Mutations. Mutations were defined as differences from the HIV-1 protease consensus B sequence (15). Of 2,244 sequences meeting the study criteria, 89% (1,990) were determined by direct PCR (population-based) sequencing and 11% (254) were determined by sequencing multiple clones of an isolate. About 1% of nucleotide positions in the sequences determined by direct PCR sequencing contained nucleotide mixtures (defined as the presence of a second electrophoretic peak of at least 20 to 30% of the primary peak). Positions with mixtures were scored as mutations in our analysis of mutation prevalence. However, because it is not possible to determine if these mutations were present in the same genome as other mutations in the sequence, mutations present as mixtures were excluded from our covariation analysis.

For the 254 isolates for which multiple clones were sequenced, we restricted our analysis to the clone that occurred with the highest frequency. This restriction caused us to exclude 128 mutations that were present in 30% or more of the clones from an individual (but that were not present in the clone with the highest prevalence) and to include 15 mutations that, although present in the most prevalent clone, existed in <30% of the total. This restriction was necessary to prevent the inclusion of mutations from different genomes in our covariation analysis. It did not significantly change the results of our analysis of mutation prevalence.

Statistical analysis. (i) Mutation prevalence. We performed chi-square tests of independence to determine if there was an association between drug treatment and a mutation at each protease position. The chi-square statistic was based on a 2-by-2 contingency table containing the numbers of isolates from treated and untreated persons and the numbers of isolates with and without mutations.

To investigate whether there was a linear relationship between the number of protease inhibitors received and the prevalence of a mutation, we performed a logistic regression analysis in which the number of drugs was the independent variable and the presence or absence of mutation was the dependent variable. Persons were categorized in one of four groups according to the extent of treatment: one, two, three, and four or more protease inhibitors. Untreated persons were not included in this analysis.

For the chi-square and logistic regression analyses, we used the method of Benjamini and Hochberg to identify results that were statistically significant in the presence of multiple-hypothesis testing (1). This method was developed for the problem of multiple-hypothesis testing when multiple significant findings are not unexpected. As opposed to the Bonferroni correction, which divides the significance cutoff by the number of hypotheses tested (n), the Benjamini-Hochberg method ranks the hypotheses by their P values, and each hypothesis of rank r is compared with a significance cutoff, now called a false-discovery rate (FDR), divided by $(n - r)$. In this study, FDRs of 0.01 and 0.05 were used to determine statistical significance.

(ii) Mutation covariation. We investigated covariation between positions by calculating the binomial (ϕ) correlation coefficient for the simultaneous presence of mutations at two positions in the same isolate. The correlation coefficients were computed separately for the subsets of protease inhibitor-treated and untreated individuals. Statistically significant correlations were those with P values of ≤ 0.05 using a Bonferroni correction for 2,080 (i.e., the binomial coefficient $\binom{65}{2}$) pairwise comparisons. We further investigated the relationships among positions by performing a principal-components analysis (PCA) of positions found in the analysis described above to be mutated in treated persons. The matrix of binomial correlation coefficients was used as a measure of similarity between positions.

Mutational clusters were defined as clusters of three or more positions in which each member of the cluster was significantly correlated with the presence of each of the other members of the cluster (referred to as cliques in graph theory). Mutational clusters were identified by an exhaustive search technique that evaluated all possible clusters that could be formed from the statistically significant pairs of covarying residues.

(iii) Structural analysis. We used two published X-ray crystallographic structures (1hsg [3] and 1hhp [27]) and one molecular-dynamics simulation (23) of wild-type HIV-1 protease to examine the interresidue distances between positions with statistically significant frequencies of covariation. One X-ray crystallographic structure (1hsg) was of HIV-1 protease bound to indinavir, and one (1hhp) was of an unliganded enzyme. The molecular-dynamics simulation, based

TABLE 1. HIV-1 isolates and protease inhibitor exposure^a

No. of PIs	No. of isolates	No. of persons treated with:					
		APV	IDV	LPV	NFV	RTV	SQV
0	1,004	0	0	0	0	0	0
1	637	25	232	0	175	61	144
2	180	7	112	3	100	47	91
3	234	30	184	1	116	161	210
4-6	189	53	182	5	177	187	183
Total	2,244	115	710	8	568	456	628

^a PI, protease inhibitor; APV, amprenavir; IDV, indinavir; LPV, lopinavir; NFV, nelfinavir; RTV, ritonavir; SQV, saquinavir.

on the 1hhp structure, showed the flaps of the protease curled inward. The distance between two residues was considered to be the shortest interatomic distance between any atoms in the two residues. Interresidue distances were calculated between all positions in the protease dimer in each of the three structures. Residues within 8 Å of each other in at least one structure were considered to be neighboring pairs in the folded enzyme. This distance was chosen as a conservative maximum distance at which two residues may interact.

When covariation could not be explained by the proximity of the two residues, we investigated the possibility that covariation resulted from the presence of one or more linking residues, a phenomenon called chained covariation (A. S. Lapedes, B. G. Giraud, L. C. Liu, and G. D. Stormo, presented at the AMS/SIAM Conference on Statistics in Molecular Biology, Seattle, Wash., 1997). To identify chained covariation, we performed a Markov chain analysis of the statistically significant pairs of covarying residues. This analysis finds the shortest chain between a residue pair, where the chain consists entirely of correlated residues within eight Å of one another. We then counted the number of covarying pairs that could be explained by a chain of one, two, three, or more linking residues. To determine whether such chains were statistically significant, we performed a stepwise permutation analysis in which we randomly generated pairs of residues and determined whether these residues could also be linked by a chain consisting entirely of correlated, neighboring residues. Repeated permutations provided the expected number and distribution of chains of one, two, three, or more linking residues in a molecule having the size and topology of HIV-1 protease.

Nucleotide sequence accession numbers. The nucleotide sequences, mutations, drug treatment histories, and GenBank accession numbers can be downloaded as a PDF file from <http://hivdb.stanford.edu/data/pr1.html>. Of the 599 previously unpublished isolates sequenced at the Stanford University Hospital between 1 July 1997 and 31 December 2001, 383 had already been submitted to GenBank for a study of HIV subtypes in northern California (8); 216 new sequences were submitted to GenBank with this report (AF544406 to AF544621).

RESULTS

Protease inhibitor treatments. Sequences of 2,244 protease isolates from 1,919 persons met the study selection criteria. Two isolates, one before and one after a protease inhibitor was received, were included from each of 325 persons. The sequences of 1,645 isolates from 1,344 individuals were published previously; the sequences of 599 isolates from 575 individuals have not been published previously. Table 1 groups the isolates in the study according to the protease inhibitor treatments of the persons from whom isolates were obtained. Indinavir, saquinavir, and nelfinavir were each received by >500 persons. Ritonavir was received by 456 persons, ~60% of whom were receiving ritonavir at a low dose as part of a dual protease inhibitor combination. One hundred fifteen persons received amprenavir, which was approved in 1999, and eight persons received lopinavir, which was approved in 2001.

Protease mutations and their association with treatment.

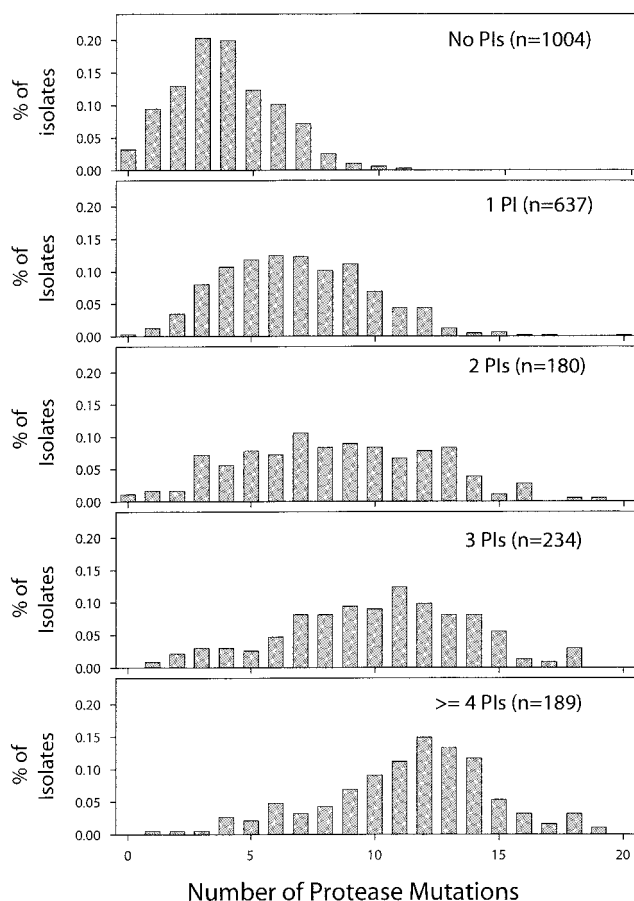


FIG. 1. Histograms of mutation frequency according to the number of protease inhibitors (PIs) received. The median number of mutations (differences from the consensus B sequence) increased from 4 in untreated persons to 12 in persons receiving ≥ 4 inhibitors.

The median number of protease mutations per isolate increased in proportion to the number of protease inhibitors received, from 4 mutations per isolate in untreated persons to 12 mutations per isolate in persons receiving four or more inhibitors (Fig. 1). Table 2 shows the mutation frequencies of the 99 protease positions according to the number of protease inhibitors received. Based on our chi-square analysis, mutations at 45 positions were found to be treatment associated in that mutation frequencies were significantly associated with treatment with at least one protease inhibitor. An additional 17 positions had non-treatment-related polymorphisms; these positions had mutations, but the mutation frequencies were not statistically associated with protease inhibitor treatment. The remaining 37 positions had mutation frequencies of $<0.5\%$, even in isolates exposed to treatment, and were considered invariant.

The 45 treatment-associated positions included 23 positions previously associated with drug resistance (10, 20, 24, 30, 32, 33, 36, 46, 47, 48, 50, 53, 54, 60, 63, 71, 73, 77, 82, 84, 88, 90, 93) and 22 positions which had not previously been associated with drug resistance (11, 13, 22, 23, 34, 35, 43, 45, 55, 58, 62, 66, 72, 74, 75, 76, 79, 83, 85, 89, 92, 95). Thirteen of the 22 newly described treatment-associated positions (positions 11, 22, 23,

45, 58, 66, 74, 75, 76, 79, 83, 85, 95) showed little or no variation—mutation frequencies of $<0.5\%$ —in untreated persons, as shown in Table 2, column 0. These 13 positions played a significant role in HIV-1 protease variation, with mutations occurring in 92 of 637 (14.4%) persons receiving a single inhibitor and 162 of 603 (26.9%) persons receiving two or more inhibitors. These mutations usually occurred in isolates with one or more primary protease inhibitor resistance mutations (219 of 254 [85.8%]).

Our logistic regression analysis revealed that mutations at 24 positions had statistically significant positive linear relationships between the number of protease inhibitors received and the presence of a mutation (Table 2). The positions with the strongest linear relationships were positions 10, 20, 46, 53, 54, 63, 71, 73, 82, 84, and 90. There was a statistically significant negative linear relationship between the number of inhibitors and the presence of a mutation at position 30.

Locations of protease mutations within the enzyme's three-dimensional structure. The invariant HIV-1 protease positions include the active-site positions (positions 25 to 27); other positions in or near the substrate cleft (positions 28 to 29, 31, and 80 to 81); most of the N- and C-terminal domains, which together with the active site make up the dimer interface; and other positions that appear to be associated with maintaining the enzyme's conformation and flexibility (e.g., 10 conserved glycines, including 3 in the flexible tips of the enzyme flap at positions 49, 51, and 52). The polymorphic positions are found almost entirely in surface loops.

The 23 known drug resistance positions include six substrate cleft residues (positions 30, 32, 48, 50, 82, and 84); four flap tip drug resistance mutations (positions 46, 47, 53, and 54); position 90, which although not in the substrate cleft decreases susceptibility to multiple protease inhibitors; three additional residues which are generally mutated only in treated persons (positions 24, 73, and 88); and nine polymorphic residues (positions 10, 20, 33, 36, 60, 63, 71, 77, and 93). The 22 new drug resistance positions include one substrate cleft residue (position 23), three flap residues (positions 43, 45, 55), one terminal-domain residue (position 95), and 17 residues in the enzyme core. The substrate cleft residues at positions 48 and 50 are also in the protease flap tips.

Correlations between protease mutations. To identify patterns of drug resistance mutations, we calculated the pairwise binary (ϕ) correlation coefficients among the 45 treatment-associated and 17 polymorphic protease residues. This analysis was performed separately for the 1,004 isolates from untreated persons and for the 1,240 isolates from treated persons to detect associations that were independent of the treatment status of the individuals from whom the sequenced isolates were obtained. Among the untreated isolates, 23 of the 2,080 possible pairwise correlations were statistically significant, including 19 positive ($\phi = 0.14$ to 0.31) and 4 negative ($\phi = -0.14$ to -0.21) correlations. Among the treated isolates, 115 of the possible 2,080 correlations were statistically significant, including 99 positive ($\phi = 0.13$ to 0.63) and 16 negative ($\phi = -0.13$ to -0.34) correlations.

Table 3 shows the most strongly correlated pairs of positions among the 115 statistically significant correlations in isolates from treated persons. The three most strongly correlated pairs of positions among the treated isolates were 54 and 82 ($\phi =$

TABLE 2. Mutation frequencies at protease positions 1 to 99 according to the number of protease inhibitors received

Position ^a	aa ^b	Mutation frequency at no. of PIs ^c					Association with PIs (P) ^d	Association with no. of PIs ^e		Common substitutions ^f
		0	1	2	3	≥4		Coefficient	P	
1	P	0.0	0.0	0.0	0.0	0.0				
2	Q	0.2	0.2	0.6	0.0	0.0				
3	I	0.4	0.0	2.2	0.0	0.0				
4	T	0.1	0.0	0.6	0.4	0.5				
5	L	0.0	0.5	0.6	0.4	0.0				
6	W	0.1	0.3	0.6	0.9	0.5				
7	Q	0.2	0.2	0.0	0.9	0.0				
8	R	0.3	0.2	0.0	0.4	0.0				
9	P	0.2	0.0	0.0	0.0	0.5				
10	L	11.6	33.8	52.2	69.2	77.2	<1.0e-9 ^g	0.66	<1.0e-9	IVFR
11 ^h	V	0.2	0.6	1.1	3.0	3.7	1.6e-3	0.60	4.7e-4	I
12	T	11.6	8.3	7.2	6.0	10.6	6.6e-3	0.02	7.6e-1	SPAKEI
13 ^h	I	12.5	21.0	23.9	15.8	15.9	8.0e-6	-0.11	5.2e-2	V
14	K	11.0	12.2	10.0	9.0	9.5	9.9e-1	-0.11	1.5e-1	R
15	I	15.0	18.5	14.4	21.4	21.7	1.7e-2	0.07	2.6e-1	V
16	G	2.6	2.8	2.2	2.1	3.7	9.3e-1	0.02	9.0e-1	EA
17	G	0.3	0.3	0.6	0.4	0.5				
18	Q	2.0	1.7	1.7	1.7	3.2	9.5e-1	0.19	2.5e-1	H
19	L	8.6	9.7	13.3	9.4	13.2	1.0e-1	0.08	2.8e-1	IVTQ
20	K	1.6	12.1	19.4	27.8	29.6	<1.0e-9	0.37	<1.0e-9	RIMT
21	E	0.2	0.0	0.0	0.0	0.0				
22 ^h	A	0.0	0.6	0.0	1.7	1.6	7.2e-3 ⁱ	0.33	1.6e-1	V
23 ^h	L	0.1	0.6	2.2	3.8	1.6	5.0e-4	0.34	5.0e-1	I
24	L	0.1	5.2	5.6	10.7	9.5	<1.0e-9	0.24	7.5e-3	I
25	D	0.0	0.2	0.0	0.9	0.0				
26	T	0.0	0.0	0.0	0.0	0.0				
27	G	0.0	0.0	0.0	0.0	0.0				
28	A	0.0	0.0	0.0	0.9	0.0				
29	D	0.3	0.0	0.0	0.4	0.0				
30	D	0.0	13.8	12.2	3.4	3.2	<1.0e-9	-0.54	1.4e-8	N
31	T	0.1	0.2	0.0	0.0	0.0				
32	V	0.0	4.2	3.9	3.8	6.3	<1.0e-9	0.11	3.2e-1	I
33	L	2.9	4.7	4.4	7.3	14.8	5.9e-5	0.41	3.0e-6	FIV
34 ^h	E	0.5	0.9	2.8	3.0	4.8	1.6e-3	0.51	4.7e-4	QD
35 ^h	E	24.0	28.7	28.9	32.1	39.7	2.6e-4	0.13	1.3e-2	DG
36	M	13.3	27.8	30.6	33.8	44.4	<1.0e-9	0.20	5.6e-4	IVL
37	N	34.6	34.5	31.1	35.0	35.4	9.2e-1	0.01	7.7e-1	DSTEHCA
38	L	0.5	0.5	0.6	0.0	0.5	9.9e-1	-0.15	6.9e-1	
39	P	3.2	3.5	1.1	1.3	2.6	4.7e-1	-0.23	1.6e-1	SQ
40	G	0.1	0.2	0.0	0.0	0.0				
41	R	21.9	17.0	19.4	22.6	19.0	6.8e-1	0.07	2.6e-1	K
42	W	0.0	0.0	0.0	0.0	0.5				
43 ^h	K	2.1	2.8	7.8	7.3	7.9	2.4e-4	0.31	1.6e-3	TR
44	P	0.1	0.0	0.0	0.0	0.0				
45 ^h	K	0.4	3.0	0.6	3.8	4.2	1.1e-5	0.14	2.9e-1	KR
46	M	0.2	22.4	38.9	42.7	51.3	<1.0e-9	0.41	<1.0e-9	ILV
47	I	0.3	2.2	1.1	1.7	3.2	3.7e-4	0.10	5.3e-1	V
48	G	0.1	3.5	5.6	12.8	10.1	<1.0e-9	0.40	8.5e-6	V
49	G	0.1	0.0	0.0	0.0	0.0				
50	I	0.1	1.4	0.6	0.9	1.4	5.0e-4	0.40	2.1e-2	V
51	G	0.1	0.3	0.6	0.9	0.0				
52	G	0.1	0.2	0.0	0.0	0.0				
53	F	0.4	1.4	5.0	11.1	15.3	<1.0e-9	0.68	<1.0e-9	LY
54	I	0.3	13.2	20.6	42.7	46.0	<1.0e-9	0.60	<1.0e-9	VLTM
55 ^h	K	0.7	2.2	4.4	03.4	7.4	1.3e-5	0.36	2.4e-3	R
56	V	0.0	0.0	0.0	0.0	0.0				
57	R	7.5	8.5	8.3	8.1	15.9	1.0e-1	0.16	3.7e-2	K
58 ^h	Q	0.3	2.7	3.9	6.8	7.4	2.9e-9	0.36	9.7e-4	E
59	Y	0.3	0.2	0.0	0.0	0.0				
60	D	3.7	6.4	7.2	13.7	13.8	1.0e-6	0.28	3.8e-4	E
61	Q	3.6	2.7	6.7	5.1	5.8	5.3e-1	0.23	3.2e-2	ENH
62 ^h	I	17.2	27.8	32.8	36.3	44.4	<1.0e-9	0.21	1.4e-5	V
63	L	68.0	82.6	90.6	91.5	96.8	<1.0e-9	0.52	<1.0e-9	PSAQTCHV
64	I	21.2	25.9	18.9	22.6	21.2	2.1e-1	-0.09	9.8e-2	VLM
65	E	2.0	2.0	2.2	1.3	1.1	8.2e-1	-0.21	2.6e-1	D

Continued on following page

TABLE 2—Continued

Position ^a	aa ^b	Mutation frequency at no. of PIs ^c					Association with PIs (P) ^d	Association with no. of PIs ^e		Common substitutions ^f
		0	1	2	3	≥4		Coefficient	P	
66^h	I	0.1	1.4	3.9	2.6	3.2	1.6e-5	0.21	1.7e-1	FV
67	C	1.9	2.4	3.3	1.7	5.3	2.0e-1	0.20	1.3e-1	SFY
68	G	0.2	0.0	0.6	0.5	0.4				
69	H	5.4	5.2	7.2	6.0	11.1	2.9e-1	0.24	7.8e-3	HKYR
70	K	3.5	3.3	2.2	4.3	1.6	6.6e-1	-0.10	4.7e-1	RQ
71	A	6.8	34.4	50.0	57.3	69.3	< 1.0e-9	0.46	< 1.0e-9	VTI
72^h	I	8.3	15.2	13.9	21.8	16.4	1.1e-8	0.08	2.3e-1	VTMLE
73	G	0.0	6.1	14.4	27.4	34.9	< 1.0e-9	0.66	< 1.0e-9	STC
74^h	T	0.2	6.6	9.4	7.7	8.5	< 1.0e-9	0.07	4.5e-1	SPA
75^h	V	0.1	1.3	1.7	2.1	0.0	2.8e-3	-0.14	5.1e-1	I
76^h	L	0.1	1.3	2.2	1.3	0.5	2.8e-3	-0.14	5.1e-1	V
77	V	23.4	28.6	36.7	34.6	34.9	1.2e-5	0.11	3.0e-2	I
78	G	0.0	0.0	0.0	0.0	0.0				
79^h	P	0.0	0.5	1.7	1.7	4.2	3.2e-4	0.59	9.1e-4	A
80	T	0.0	0.0	0.0	0.0	0.0				
81	P	0.0	0.2	0.0	0.0	0.0				
82	V	1.2	25.0	33.3	46.2	42.3	< 1.0e-9	0.30	1.6e-9	ATFI
83^h	N	0.0	0.5	1.1	0.9	1.6	1.1e-2ⁱ	0.31	2.1e-1	D
84	I	0.0	5.0	16.1	24.8	36.5	< 1.0e-9	0.71	< 1.0e-9	V
85^h	I	0.2	2.8	1.7	4.7	5.7	9.3e-8	0.26	3.6e-2	V
86	G	0.0	0.2	0.0	0.0	0.0				
87	R	0.6	0.0	0.0	1.3	0.0				
88	N	0.2	9.1	8.9	6.8	7.4	< 1.0e-9	-0.09	3.2e-1	DS
89^h	L	0.7	1.4	5.0	7.3	4.2	1.9e-5	0.41	6.3e-4	MVI
90	L	0.1	22.3	40.6	55.6	66.7	< 1.0e-9	0.64	< 1.0e-9	M
91	T	0.6	0.2	0.0	0.4	0.0				
92^h	Q	0.9	2.4	3.9	3.8	3.2	6.3e-4	0.16	2.4e-1	KR
93	I	24.1	34.5	36.7	42.3	48.7	< 1.0e-9	0.17	3.0e-4	LM
94	G	0.0	0.0	0.0	0.0	0.0				
95^h	C	0.3	0.6	2.2	3.8	3.2	1.3e-3	0.47	3.1e-3	F
96	T	0.0	0.0	0.0	0.0	0.0				
97	L	0.1	0.0	0.0	0.0	0.0				
98	N	0.3	0.2	0.0	0.0	0.0				
99	F	0.0	0.0	0.6	0.0	0.0				

^a Positions that were associated with PIs are in boldface.

^b aa, amino acid.

^c PIs, protease inhibitors.

^d Chi-square test of independence between presence of mutation and having received a PI.

^e Logistic regression of the occurrence of mutation on the number of PIs. P values are shown for the 62 variable positions. Coefficient, logistic regression coefficient.

^f Mutations occurring in five or more isolates listed in the order of frequency.

^g P values in bold were significant at an FDR of 0.01 following correction for multiple comparisons.

^h Newly described treatment-associated mutations.

ⁱ A22V and N83D were significant at an FDR of 0.05.

0.63), 32 and 47 (phi = 0.51), and 73 and 90 (phi = 0.47). Mutations at two pairs of primary resistance positions had significant positive correlations: positions 84 and 90 and positions 48 and 82. Mutations at positions 82 and 90, although both common, were not significantly correlated with each other. Position 30 was negatively correlated with each of the other primary resistance positions. The positions with the greatest number of positive correlations were positions 10 (16 correlations), 46 (13 correlations), 71 (12 correlations), 90 (10 correlations), 20 (10 correlations), 73 (10 correlations), 82 (9 correlations), 63 (7 correlations), 84 (6 correlations), and 54 (6 correlations).

Correlations usually involved the most common mutation at each of the two correlated positions (Table 3). For example, the strong positive correlation between positions 54 and 82 (phi = 0.63) is in large part due to the strong correlation between I54V, the most common substitution at position 54, and V82A, the most common substitution at position 82 (for

I54V and V82A, phi = 0.55). Other combinations of substitutions for these two positions were less commonly observed: I54T and V82A (phi = 0.21) and I54V and V82T (phi = 0.15). In some cases, covariation was dominated very strongly by particular combinations of substitutions. For example the positive correlation between positions 30 and 88 (phi = 0.40) was represented entirely by D30N and N88D (phi = 0.52) rather than by D30N and N88S (phi = -0.05), and the correlation between positions 48 and 54 (phi = 0.29) was represented largely by G48V and I54T (phi = 0.44) rather than G48V and I54V (phi = 0.19).

We can use our measurements of comutation frequencies to construct a graphical model that summarizes the relationships among positions in HIV-1 protease. In this model, we attempt to place positions with high degrees of comutation close together and positions with low or negative degrees of comutation far apart. These relationships are modeled as consistently as possible within the framework of a two-dimensional plot.

TABLE 3. Most strongly correlated pairs of positions among 115 statistically significant correlations in isolates from treated persons^a

Pos 1	Pos 2	Residue role(s) ^b	Phi	P value ^c	(Å) Distance ^d	Correlations between specific amino acids			
						Pos 1	Pos 2	Phi	P value ^c
Positive correlations									
82	54	Substrate cleft/flap	0.63	<1.0e-9	5.6	82A	54V	0.55	<1.0e-9
						82A	54T	0.21	<1.0e-9
						82T	54V	0.15	7.5e-7
						82I	54M	0.15	1.4e-4
32	47	Substrate cleft/flap	0.51	<1.0e-9	3.9	32I	47V	0.51	<1.0e-9
90	73	90M/accessory	0.47	<1.0e-9	8.9	90M	73S	0.38	<1.0e-9
						90M	73T	0.18	1.0e-9
36	35	Accessory/polymorphism	0.45	<1.0e-9	1.3	36I	35D	0.45	<1.0e-9
54	10	Flap/accessory	0.41	<1.0e-9	16.3	54V	10I	0.38	<1.0e-9
20	36	Accessory/accessory	0.41	<1.0e-9	3.2	20R	36I	0.40	<1.0e-9
30	88	Substrate cleft/accessory	0.40	<1.0e-9	3.5	30N	88D	0.52	<1.0e-9
90	71	90M/accessory	0.38	<1.0e-9	6.3	90M	71V	0.30	<1.0e-9
10	71	Accessory/accessory	0.37	<1.0e-9	15.1	10I	71V	0.32	<1.0e-9
90	10	90M/accessory	0.35	<1.0e-9	10.1	90M	10I	0.27	<1.0e-9
82	10	Substrate cleft/accessory	0.35	<1.0e-9	4.8	82A	10I	0.32	<1.0e-9
						82T	10R	0.25	<1.0e-9
46	10	Flap/accessory	0.35	<1.0e-9	17.7	46I	10I	0.16	8.0e-8
						46L	10I	0.15	1.5e-7
54	71	Flap/accessory	0.34	<1.0e-9	20.4	54V	71V	0.32	<1.0e-9
77	93	Accessory/accessory	0.31	<1.0e-9	17.6	77I	93L	0.20	<1.0e-9
84	10	Substrate cleft/accessory	0.30	<1.0e-9	7.8	84V	10I	0.28	<1.0e-9
84	90	Substrate cleft/90M	0.30	<1.0e-9	7.6	84V	90M	0.29	<1.0e-9
48	54	Substrate cleft/flap	0.29	<1.0e-9	5.0	48V	54T	0.44	<1.0e-9
						48V	54V	0.18	1.5e-9
12	19	Polymorphism/polymorphism	0.29	<1.0e-9	4.2	12E	19I	0.26	<1.0e-9
84	73	Substrate cleft/accessory	0.28	<1.0e-9	10.5	84V	73S	0.23	<1.0e-9
46	24	Flap/accessory	0.27	<1.0e-9	18.8	46L	24I	0.20	<1.0e-9
						46I	24I	0.16	6.1e-8
10	73	Accessory/accessory	0.27	<1.0e-9	18.9	10I	73S	0.18	<1.0e-9
82	71	Substrate cleft/accessory	0.26	<1.0e-9	16.5	82A	71V	0.27	<1.0e-9
82	24	Substrate cleft/accessory	0.26	<1.0e-9	7.5	82A	24I	0.28	<1.0e-9
35	37	Polymorphism/polymorphism	0.26	<1.0e-9	4.6	35D	37D	0.18	<1.0e-9
						35D	37E	0.17	5.2e-9
54	24	Flap/accessory	0.26	<1.0e-9	15.8	24I	54V	0.29	<1.0e-9
36	62	Accessory/polymorphism	0.26	<1.0e-9	7.3	36I	62V	0.21	<1.0e-9
82	48	Substrate cleft/substrate cleft	0.25	<1.0e-9	7.6	82A	48V	0.27	<1.0e-9
84	71	Substrate cleft/accessory	0.25	<1.0e-9	11.0	84V	71V	0.23	<1.0e-9
82	46	Substrate cleft/flap	0.25	<1.0e-9	14.1	82A	46L	0.28	<1.0e-9
90	63	90M/accessory	0.23	<1.0e-9	11.4	90M	63P	0.31	<1.0e-9
90	20	90M/accessory	0.22	<1.0e-9	12.9	90M	20I	0.19	<1.0e-9
90	93	90M/accessory	0.22	<1.0e-9	3.3	90M	93L	0.21	<1.0e-9
10	93	Accessory/accessory	0.22	<1.0e-9	10.6	10I	93L	0.20	<1.0e-9
46	55	Flap/polymorphism	0.22	<1.0e-9	3.8	46I	55R	0.15	4.7e-7
12	14	Polymorphism/polymorphism	0.22	<1.0e-9	4.4	12P	14R	0.31	<1.0e-9
10	24	Accessory/accessory	0.22	<1.0e-9	5.0	10I	24I	0.22	<1.0e-9
30	75	Substrate cleft/nonpolymorphic	0.22	<1.0e-9	6.7	30N	75I	0.22	<1.0e-9
54	20	Flap/accessory	0.22	<1.0e-9	14.5	20R	54V	0.24	<1.0e-9
60	61	Accessory/polymorphism	0.21	<1.0e-9	1.3	60E	61E	0.24	<1.0e-9
48	10	Substrate cleft/accessory	0.21	<1.0e-9	13.3	10I	48V	0.22	<1.0e-9
71	73	Accessory/accessory	0.21	<1.0e-9	4.1	71V	73A	0.27	<1.0e-9
						71I	73T	0.24	<1.0e-9
73	62	Accessory/polymorphism	0.20	<1.0e-9	2.9	73S	62V	0.19	<1.0e-9
84	79	Substrate cleft/Nonpolymorphic	0.20	<1.0e-9	8.2	84V	79A	0.16	3.4e-7
Negative correlations									
36	77	Accessory/accessory	-0.34	<1.0e-9	3.4	36I	77I	-0.31	<1.0e-9
71	64	Accessory/polymorphism	-0.23	<1.0e-9	2.9	71V	64V	-0.18	1.0e-9
63	64	Accessory/polymorphism	-0.22	<1.0e-9	1.3	63P	64V	-0.25	<1.0e-9
30	82	Substrate cleft/substrate cleft	-0.22	<1.0e-9	8.8	30N	82A	-0.17	2.4e-9
30	90	Substrate cleft/90M	-0.21	<1.0e-9	7.9	30N	90M	-0.20	<1.0e-9
30	10	Substrate cleft/accessory	-0.20	<1.0e-9	13.3	30N	10I	-0.18	<1.0e-9
90	24	90M/accessory	-0.20	<1.0e-9	3.8	90M	24I	-0.19	<1.0e-9

^a The 43 positive correlations with phi of ≥ 0.20 and the 7 negative correlations with phi of ≤ -0.20 . Correlations between specific amino acids at these positions are shown for those cases where phi ≥ 0.15 or phi ≤ -0.15 . Pos, position.

^b Substrate cleft residues are directly in contact with one or more inhibitors. Residues in the flap and at position L90M often directly confer or contribute to protease inhibitor resistance even though they are not in direct inhibitor contact. Accessory mutations contribute to resistance only when present with a mutation in the substrate cleft or flap or at position 90. Of the polymorphisms in this table, those at positions 35, 55, and 62 are significantly associated with drug therapy, whereas those at positions 12, 14, 19, 37, and 64 are not associated with drug therapy (Table 2). The nonpolymorphic residues at positions 75 and 79 are significantly associated with drug therapy (Table 2). The substrate cleft residues at positions 48 and 50 are also in the flap.

^c P values shown have not been corrected for multiple comparisons.

^d The distance between two residues was considered to be the shortest interatomic distance between each of the atoms in the two residues. Residues 82 and 54 are 8.3 Å apart in the 1hsg structure (3) but only 5.6 Å apart in the molecular-dynamics simulation, both described in Materials and Methods (23). The remaining distances are based on the 1hsg structure.

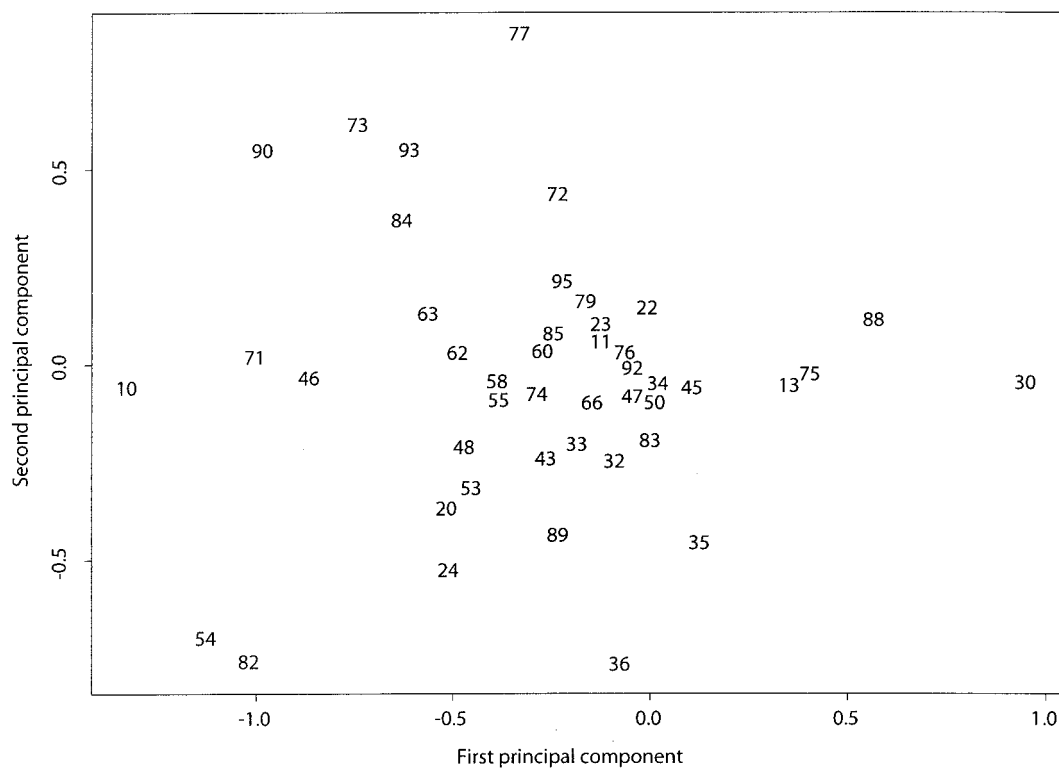


FIG. 2. PCA of the 45 positions associated with protease inhibitor treatment. The graph is a two-dimensional projection of the distances among the 45 positions, where the similarity between any two positions is measured by their binary (ϕ) correlation coefficient among persons who have received at least one inhibitor. Positions with high degrees of comutation are close together, and positions with low or negative degrees of comutation are far apart. These relationships are modeled as consistently as possible within the framework of a two-dimensional plot.

One computational technique that generates such graphical models is called PCA. We performed PCA on the 45 positions that were associated with protease inhibitor treatment and used the matrix of correlation coefficients as a measure of similarity between positions. The results of our PCA are shown in Fig. 2. The figure shows that positions 30 and 88 cluster together and are separate from most other positions. It also shows a clustering of positions 54 and 82 and their separation from positions 73, 84, 90, and 93.

Correlated mutations and protease residue contacts. Among the 115 correlated residue pairs, 59 (51%) contained residues that were within 8 Å of each other—many more than the 5.5 pairs predicted when 115 pairs were selected at random. Most of the 59 pairs were close in each of the three structures we examined (liganded, unliganded, and open flap), but four were close only in the open-flap structure from a molecular-dynamics simulation. For example, residues 54 and 82 were separated by 5.4 Å in the open-flap structure but by 8.4 and 8.6 Å in the liganded and unliganded structures, respectively. One of the residue pairs could be explained only by contact between residues on different chains of the protease dimer (residue pair 48 and 82).

Fifty-six (49%) of the 115 correlated pairs were separated by >8 Å. Our Markov chain analysis showed that of these 56 pairs of residues, 16 could be linked by one residue, 21 by two residues, 13 by three residues, and 1 by five residues. However, our permutation analysis, which was designed to determine whether such chains were statistically significant, showed that

this amount of chained covariation would be expected by chance in a molecule with 56 correlated, neighboring residues having the size and topology of HIV-1 protease. Therefore, compared with randomly selected residue pairs, the covarying residues we observed were significantly more likely to be within 8 Å of one another but not significantly more likely to be linked by chained covariation.

Figure 3 shows the strongest positive correlations superimposed on the structure of the protease. Most of the strong correlations are in a plane that is adjacent to the substrate cleft and include residues 10, 24, 30, 46, 54, 82, 84, and 90.

Clusters of correlated residues. Pairs of correlated residues can be further grouped into clusters in which all possible pairs within the cluster are mutually correlated. Among the 23 highly correlated pairs found in isolates from untreated persons, there were two mutational clusters, one of three residues and one of four residues. Among the 115 correlated pairs found in isolates from treated persons, there were 30 mutational clusters, ranging in size from three to six residues (Table 4).

Twenty of the 30 clusters in treated isolates contained one or more primary protease mutations, including L90M (12 clusters), V82ATF (6 clusters), and D30N (6 clusters). The substrate cleft mutation I84V was in four of the L90M clusters. The substrate cleft mutation G48V was in one of the V82ATF clusters. Flap tip positions were included in 4 of the 12 clusters containing L90M (position 46, 4 clusters), and each of the 6 clusters containing V82ATF (position 46, 4 clusters; position 53, 1 cluster; position 54, 4 clusters).

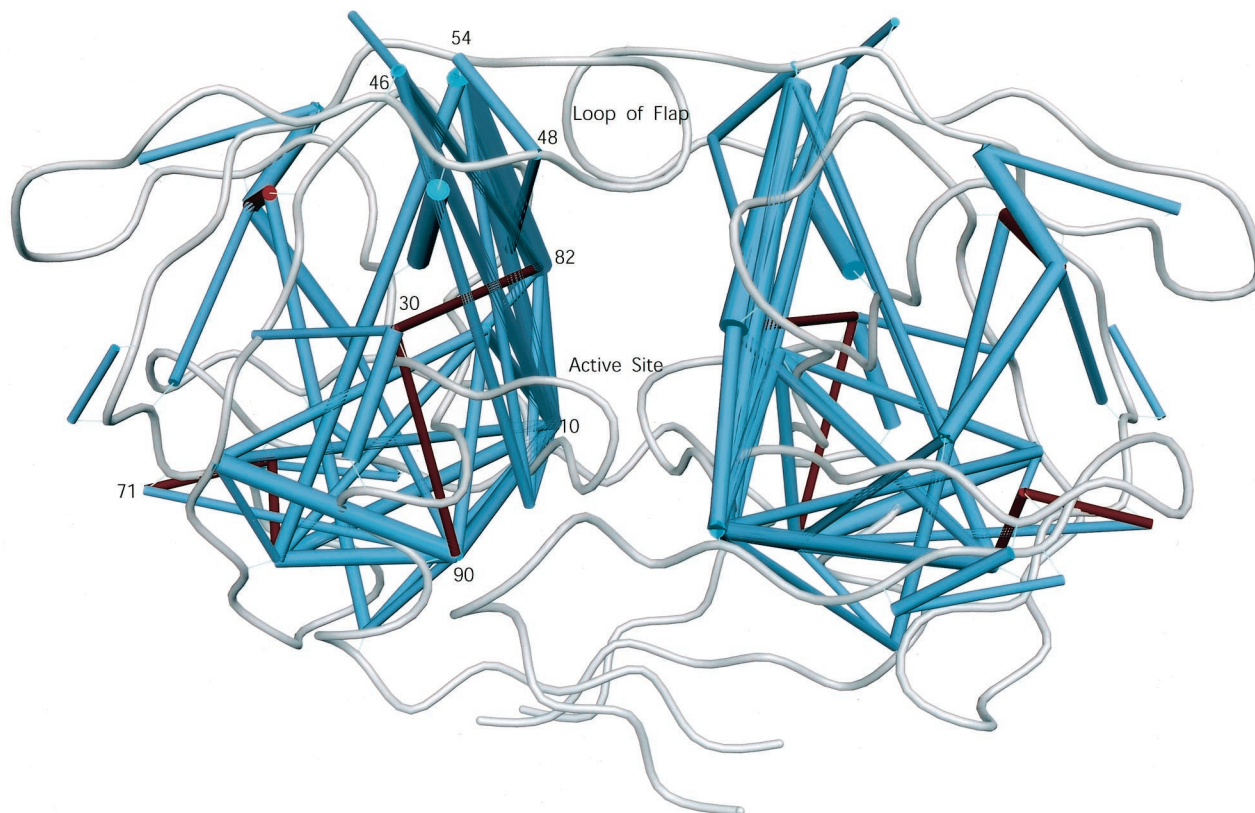


FIG. 3. The 50 most highly correlated residues in isolates from treated persons are shown superimposed on the locations of these residues within the folded enzyme. The blue lines represent positively correlated residues ($n = 44$; $\phi > 0.2$); the red lines represent negatively correlated residues ($n = 7$; $\phi < -0.2$). The diameter of each line is proportional to the correlation coefficient of the residue pair. The lines connect the beta carbons of each residue, with the exception of the glycines at positions 48 and 73, which are connected to other residues by their alpha carbons. Each correlated pair is shown twice, once in each monomer.

Six representative clusters from Table 4 are shown in Fig. 4. These six clusters occurred in 17% of isolates from all treated persons and 29% of isolates from persons receiving two or more protease inhibitors. Published *in vitro* susceptibility results for isolates containing each of these six patterns of mutations (and no additional known resistance mutations) reveal that each pattern is associated with reduced susceptibility to each of the protease inhibitors: amprenavir, 2- to 5-fold; indinavir, 10- to 15-fold; lopinavir, 2- to 20-fold; nelfinavir, 10- to 30-fold; saquinavir, 3- to 30-fold; and ritonavir, 3- to 100-fold (25).

DISCUSSION

Mutation prevalence. Of the 99 amino acids in HIV-1 protease, we found that 45 exhibit treatment-associated mutations, 17 have non-treatment-related polymorphisms, and 37 rarely if ever vary. Only subtype B isolates were included in this analysis because few sequences of non-B isolates from persons receiving antiretroviral therapy are available. However, although each subtype is characterized by different polymorphisms, the 37 invariant positions in subtype B isolates are also highly conserved in non-B protease isolates (<1% mutation frequency) (8).

The large number of isolates analyzed in this study and the

fact that a large proportion were from patients who received multiple protease inhibitors allowed us to identify 22 new treatment-associated positions. Mutations at eight of these new positions were observed to develop in a previous longitudinal study of protease isolates from 178 treated persons, but the associations of these mutations with treatment in that study were in most cases not statistically significant (24). The newly identified mutations occur primarily in combination with previously reported drug resistance mutations, suggesting that they act as accessory mutations to increase the level of resistance to multiple protease inhibitors or to compensate for losses in fitness. Most of the new mutations involve the replacement of one hydrophobic residue with another, possibly resulting in the repacking of hydrophobic regions in the core domain of the monomer.

Of the newly identified sites of mutation, residue 23—located at the base of the P1 pocket, where it is flanked by V82 and I84—is the position most likely to have a direct impact on inhibitor binding. The mutation L23I likely tightens or reshapes the P1 pocket and may compensate for the increase in size of the pocket that occurs with either V82A or I84V. Alternatively, L23I may directly interfere with inhibitor binding, as it is near the active site (30). Site-directed mutagenesis experiments in which L23I is placed in a wild-type enzyme or

TABLE 4. Clusters of correlated protease positions

Primary mutation	Other mutations	Cluster ^a	No. (%) of isolates		
Treated isolates					
L90M	L10I ± A71VT ± G73S	10 63 71 90 93	118 (9.5)		
		10 63 71 73 90	82 (6.6)		
		10 62 63 90 93	75 (6.0)		
		10 62 63 73 90	62 (5.0)		
		10 20 71 73 90	24 (1.9)		
		10 20 62 73 90	20 (1.6)		
		10 46 71 90 93	56 (4.5)		
		10 (30) 73 84 90	53 (4.3)		
		10 (30) 46 84 90	47 (3.8)		
		10 71 73 84 90	42 (3.4)		
		10 46 71 84 90	34 (2.7)		
		10 24 46; 10 46 90 ^b	52 (4.2); 142 (11.4) ^b		
		10 (30) 46 54 82	82 (6.6)		
		10 48 54 82	40 (3.2)		
		10 24 46 54 82	32 (2.6)		
V82A	M46I ± I54V	32 46 82	23 (1.9)		
		10 46 53 54 71 82	6 (0.5)		
		30 (82) 88	44 (3.5)		
		13 30 88	18 (1.5)		
		30 75 88	3 (0.2)		
D30N	N88D	10 46 63 71 93	82 (6.6)		
		20 36 54	75 (6.0)		
		10 20 54 71	57 (4.6)		
		63 (64) 71	417 (33.6)		
		10 77 93	134 (10.8)		
		20 36 62	73 (5.8)		
		20 35 36 (77)	70 (5.6)		
		15 20 36 (77)	41 (3.3)		
		10 24 89	11 (0.9)		
		10 20 73; 10 73 77 ^c	36 (2.9); 69 (5.6) ^c		
None	M46I ± I54V	10 46 63 71 93	82 (6.6)		
		20 36 54	75 (6.0)		
None	Miscellaneous	10 20 54 71	57 (4.6)		
		63 (64) 71	417 (33.6)		
		10 77 93	134 (10.8)		
		20 36 62	73 (5.8)		
		20 35 36 (77)	70 (5.6)		
		15 20 36 (77)	41 (3.3)		
		10 24 89	11 (0.9)		
		10 20 73; 10 73 77 ^c	36 (2.9); 69 (5.6) ^c		
		Untreated isolates			
		None	A71VT + I93L	71, 77, 93	25 (0.2)
62, 63, 71, 93	22 (0.2)				

^a Clusters in boldface are shown in Fig. 4. Residues within parentheses are negatively correlated with the other residues in the cluster.

^b 10 + 46 + 90 occurred in 142 isolates; 10 + 24 + 46 occurred in 52 isolates; 24 and 90 were negatively correlated.

^c 10 + 73 + 77 occurred in 69 isolates; 10 + 20 + 73 occurred in 36 isolates; 20 and 77 were negatively correlated.

in an enzyme containing other mutations (e.g., V82A or I84V) are needed to clarify the effect of this mutation on protease function and protease inhibitor resistance.

Our analysis of the association between mutation prevalence and drug therapy has several limitations. First, the lack of available data on the isolates used in this study made it impossible to demonstrate a direct association between mutations and reduced in vitro susceptibility. Second, we did not control for the duration of HIV-1 infection or protease inhibitor treatment. However, despite these limitations, our analyses do establish a conservative lower limit to the extent of HIV-1 protease mutability and generate hypotheses about specific mutations. These hypotheses can be confirmed by demonstrating the longitudinal development of the mutations with treatment or the effects of the mutations on in vitro drug susceptibility.

Of the 22 newly described treatment-associated mutations, the 13 that are conserved in untreated persons are of more interest than the 9 polymorphic positions. Indeed, we cannot exclude the possibility that the increased prevalence of mutations at the nine polymorphic mutations reflects the increased variability of virus populations in persons infected for a longer

period of time—a population that is likely to include more treated than untreated persons.

A recent computational study evaluated the variability of protease residues in HIV-1, other primate lentiviruses, and feline immunodeficiency virus, as well as the theoretical free-energy contribution of each residue to the binding of HIV-1 substrates and inhibitors (29). Our analyses complement this effort by quantifying the variability of this enzyme in isolates that have evolved in the presence of one or more protease inhibitors. However, positions reported to be invariant may develop mutations within virus populations under different selection pressures. Mutations other than those described in this paper have been reported during in vitro passage experiments. Mutation at the invariant residue 91 (T91S) has been reported during in vitro passage with lopinavir (2). The substrate cleft mutations R8QK and A28S have been reported after passage with the experimental inhibitors A-77003 and TMC-126, respectively (11, 31).

Mutation covariation. The presence of positions within a molecule that covary, or mutate in a correlated manner, suggests that mutations at one position may require a compensatory mutation at a second position for optimal function (7, 16;

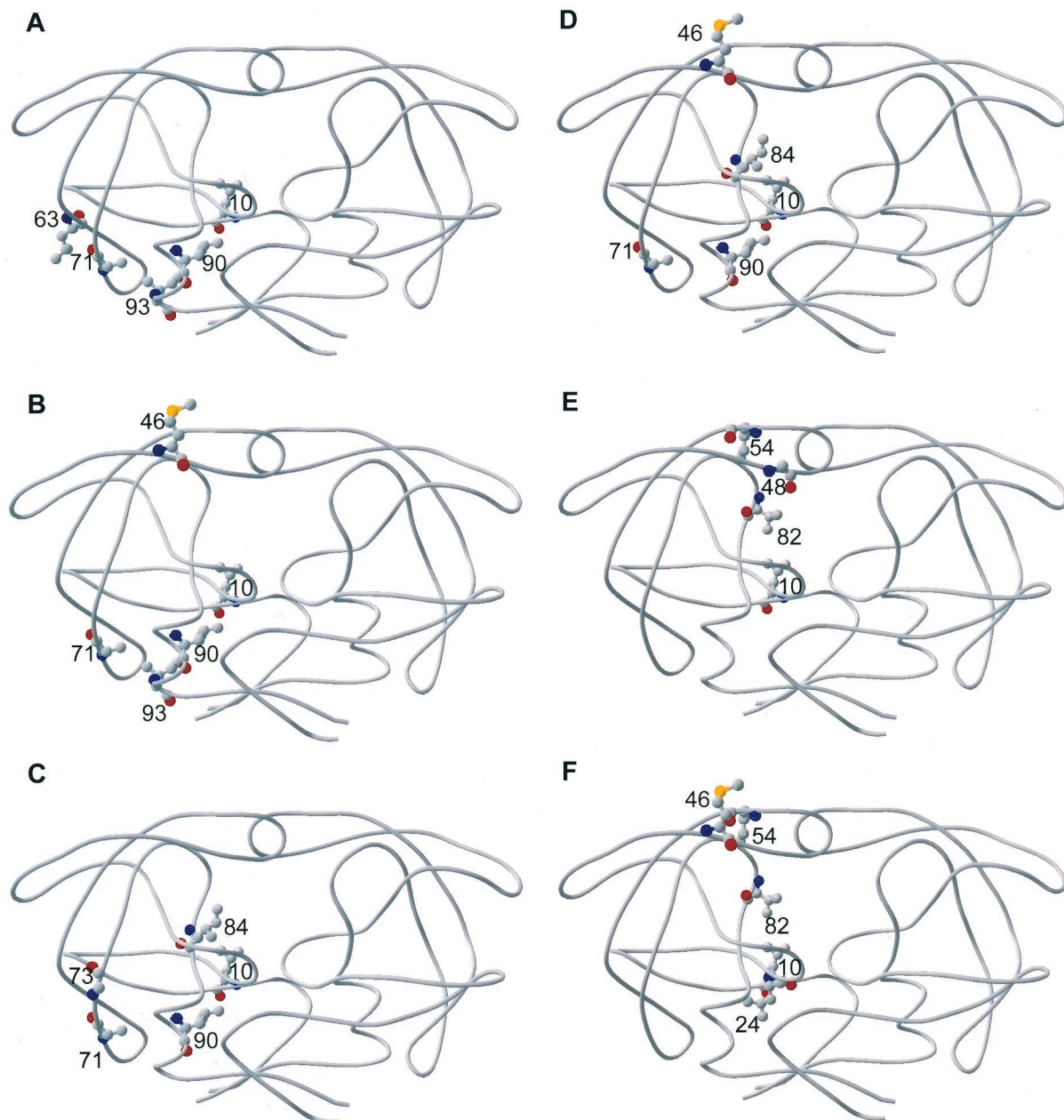


FIG. 4. Six representative clusters from Table 4. Each position in a cluster demonstrates statistically significant mutational covariation with each of the other positions within a cluster. (A) Positions 10, 63, 71, 90, and 93; (B) positions 10, 46, 71, 90, and 93; (C) positions 10, 71, 73, 84, and 90; (D) positions 10, 46, 71, 84, and 90; (E) positions 10, 48, 54, and 82; (F) positions 10, 24, 46, 54, and 82. The clusters are only shown on one monomer of the protease dimer. The side chains of the residues within each cluster are shown on the protease backbone. Oxygen is shown in red, nitrogen in blue, carbon in gray, and sulfur in yellow.

Lapedes et al., AMS/SIAM Conference on Statistics in Molecular Biology). Covariation analysis has been used to help predict unsolved protein structures and to better understand the functions of proteins with known structures. Previous analyses of covariation have used alignments of sequences in a protein family rather than an alignment of variants of a single protein. However, the high mutation rate and mutation tolerance of HIV-1 have made it possible for us and others (14) to identify

statistically significant covariation within a single HIV-1 subtype.

One of the major challenges of covariation analysis is to differentiate covariation resulting from the functional dependency between two positions from the shared inheritance of both mutations from a founder virus. In this study, covariation almost certainly reflects functionality rather than evolutionary relatedness. The fact that mutational correlations were so

much more common among treated isolates is consistent with the repeated selection of the correlated mutations in many different isolates during selective drug pressure rather than with the inheritance of the correlated mutations from a small number of ancestral isolates.

Although biochemical and biophysical experiments are required to demonstrate the mechanism for the correlation between pairs of residues, our analyses provide preliminary hypotheses that help prioritize which residues to study. For example, more than one-half of the 115 pairs of significantly correlated positions were within 8 Å of each other. This proportion exceeds that expected by chance, suggesting that in many cases covariation results from a direct interaction between the correlated mutations.

The correlations between amino acids that were not close to one another in the three-dimensional protease structure are more difficult to explain. For example, mutations at position 46 were highly correlated with mutations at many distant positions (Fig. 3 and 4). Although we found that many nonneighboring but highly correlated residues could be linked through a chain of covarying residues, our statistical analysis suggested that the compact shape of the protease and the large number of correlated residues could cause this to occur by chance. An alternative explanation for correlation between distant residues comes from the work of others who have shown that the wild-type protease enzyme may be partially down-regulated and that mutations at certain residues, such as M46I and L63P, increase catalytic activity and may be selected in enzymes with other mutations that decrease catalytic activity, regardless of the locations of these other mutations (9, 22).

The negative correlation between D30N and the other primary protease inhibitor resistance mutations may reflect the fact that D30N decreases protease fitness without contributing resistance to any protease inhibitor other than nelfinavir (6, 12, 18). Alternatively, enzymes containing D30N together with other primary mutations appear to have decreased activity (28).

The frequent occurrence of mutational clusters, as well as other common patterns of mutations, suggests that mutations can interact as part of higher-order networks. These mutation patterns are tangible evidence for the high genetic barrier to resistance to the protease inhibitors. However, these patterns are complex and frequently overlapping, suggesting that there are few, if any, absolute dependencies between drug resistance mutations. Determining the biochemical and biophysical properties of enzymes with these patterns of mutations will be important for designing new protease inhibitors that are less likely to trigger resistance or are effective against already drug-resistant isolates.

ACKNOWLEDGMENTS

M.J.G. and R.W.S. were supported in part by NIH/NIAID (AI-46148-02). R.K. was supported in part by a Stanford University BioX Interdisciplinary Project Grant. R.W.S. and C.A.S. were supported in part by NIH/NIGMS P01 GM66524-01.

REFERENCES

- Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57:289–300.
- Carrillo, A., K. D. Stewart, H. L. Sham, D. W. Norbeck, W. E. Kohlbrenner, J. M. Leonard, D. J. Kempf, and A. Molla. 1998. In vitro selection and characterization of human immunodeficiency virus type 1 variants with increased resistance to ABT-378, a novel protease inhibitor. *J. Virol.* 72:7532–7541.
- Chen, Z., Y. Li, E. Chen, D. L. Hall, P. L. Darke, C. Culberson, J. A. Shafer, and L. C. Kuo. 1994. Crystal structure at 1.9-Å resolution of human immunodeficiency virus (HIV) II protease complexed with L-735,524, an orally bioavailable inhibitor of the HIV proteases. *J. Biol. Chem.* 269:26344–26348.
- Condra, J. H., D. J. Holder, W. A. Schleif, O. M. Blahy, R. M. Danovich, L. J. Gabryelski, D. J. Graham, D. Laird, J. C. Quintero, A. Rhodes, H. L. Robbins, E. Roth, M. Shivaprakash, T. Yang, J. A. Chodakewitz, P. J. Deutsch, R. Y. Leavitt, F. E. Massari, J. W. Mellors, K. E. Squires, R. T. Steigbigel, H. Teppler, and E. A. Emimi. 1996. Genetic correlates of in vivo viral resistance to indinavir, a human immunodeficiency virus type 1 protease inhibitor. *J. Virol.* 70:8270–8276.
- Condra, J. H., W. A. Schleif, O. M. Blahy, L. J. Gabryelski, D. J. Graham, J. C. Quintero, A. Rhodes, H. L. Robbins, E. Roth, and M. Shivaprakash. 1995. In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature* 374:569–571.
- Devereux, H. L., V. C. Emery, M. A. Johnson, and C. Loveday. 2001. Replicative fitness in vivo of HIV-1 variants with multiple drug resistance-associated mutations. *J. Med. Virol.* 65:218–224.
- Gobel, U., C. Sander, R. Schneider, and A. Valencia. 1994. Correlated mutations and residue contacts in proteins. *Proteins* 18:309–317.
- Gonzales, M. J., R. N. Machekano, and R. W. Shafer. 2001. Human immunodeficiency virus type 1 reverse-transcriptase and protease subtypes: classification, amino acid mutation patterns, and prevalence in a northern California clinic-based population. *J. Infect. Dis.* 184:998–1006.
- Gulnik, S. V., L. I. Suvorov, B. Liu, B. Yu, B. Anderson, H. Mitsuya, and J. W. Erickson. 1995. Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry* 34:9282–9287.
- Hertogs, K., S. Bloor, S. D. Kemp, C. Van den Eynde, T. M. Alcorn, R. Pauwels, M. Van Houtte, S. Staszewski, V. Miller, and B. A. Larder. 2000. Phenotypic and genotypic analysis of clinical HIV-1 isolates reveals extensive protease inhibitor cross-resistance: a survey of over 6000 samples. *AIDS* 14:1203–1210.
- Ho, D. D., T. Toyoshima, H. Mo, D. J. Kempf, D. Norbeck, C. M. Chen, N. E. Wideburg, S. K. Burt, J. W. Erickson, and M. K. Singh. 1994. Characterization of human immunodeficiency virus type 1 variants with increased resistance to a C2-symmetric protease inhibitor. *J. Virol.* 68:2016–2020.
- Kantor, R., W. J. Fessel, A. R. Zolopa, D. Israelski, N. Shulman, J. G. Montoya, M. Harbour, J. M. Schapiro, and R. W. Shafer. 2002. Evolution of primary protease inhibitor resistance mutations during protease inhibitor salvage therapy. *Antimicrob. Agents Chemother.* 46:1086–1092.
- Kempf, D. J., J. D. Isaacson, M. S. King, S. C. Brun, Y. Xu, K. Real, B. M. Bernstein, A. J. Japour, E. Sun, and R. A. Rode. 2001. Identification of genotypic changes in human immunodeficiency virus protease that correlate with reduced susceptibility to the protease inhibitor lopinavir among viral isolates from protease inhibitor-experienced patients. *J. Virol.* 75:7462–7469.
- Korber, B. T., R. M. Farber, D. H. Wolpert, and A. S. Lapedes. 1993. Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis. *Proc. Natl. Acad. Sci. USA* 90:7176–7180.
- Kuiken, C. L., B. Foley, B. H. Hahn, P. Marx, F. E. McCutchan, J. Mellors, J. I. Mullins, S. Wolanski, and B. Korber. 1999. Human retroviruses and AIDS: a compilation and analysis of nucleic and amino acid sequences. Los Alamos National Laboratory, Los Alamos, N.Mex.
- Larson, S. M., A. A. Di Nardo, and A. R. Davidson. 2000. Analysis of covariation in an SH3 domain sequence alignment: applications in tertiary contact prediction and the design of compensating hydrophobic core substitutions. *J. Mol. Biol.* 303:433–446.
- Maguire, M., D. Shortino, A. Klein, W. Harris, V. Manohitharajah, M. Tisdale, R. Elston, J. Yeo, S. Randall, F. Xu, H. Parker, J. May, and W. Snowden. 2002. Emergence of resistance to protease inhibitor amprenavir in human immunodeficiency virus type 1-infected patients: selection of four alternative viral protease genotypes and influence of viral susceptibility to coadministered reverse transcriptase nucleoside inhibitors. *Antimicrob. Agents Chemother.* 46:731–738.
- Martinez-Picado, J., A. V. Savara, L. Sutton, and R. T. D'Aquila. 1999. Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. *J. Virol.* 73:3744–3752.
- Molla, A., M. Korneyeva, Q. Gao, S. Vasavanonda, P. J. Schipper, H. M. Mo, M. Markowitz, T. Chernyavskiy, P. Niu, N. Lyons, A. Hsu, G. R. Granneman, D. D. Ho, C. A. Boucher, J. M. Leonard, D. W. Norbeck, and D. J. Kempf. 1996. Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. *Nat. Med.* 2:760–766.
- Patick, A. K., M. Duran, Y. Cao, D. Shugarts, M. R. Keller, E. Mazabel, M. Knowles, S. Chapman, D. R. Kuritzkes, and M. Markowitz. 1998. Genotypic and phenotypic characterization of human immunodeficiency virus type 1 variants isolated from patients treated with the protease inhibitor nelfinavir. *Antimicrob. Agents Chemother.* 42:2637–2644.

21. **Schapiro, J. M., M. A. Winters, F. Stewart, B. Efron, J. Norris, M. J. Kozal, and T. C. Merigan.** 1996. The effect of high-dose saquinavir on viral load and CD4+ T-cell counts in HIV-infected patients. *Ann. Intern. Med.* **124**:1039–1050.
22. **Schock, H. B., V. M. Garsky, and L. C. Kuo.** 1996. Mutational anatomy of an HIV-1 protease variant conferring cross-resistance to protease inhibitors in clinical trials. Compensatory modulations of binding and activity. *J. Biol. Chem.* **271**:31957–31963.
23. **Scott, W. R., and C. A. Schiffer.** 2000. Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. *Struct. Fold. Des.* **8**:1259–1265.
24. **Shafer, R. W., P. Hsu, A. K. Patick, C. Craig, and V. Brendel.** 1999. Identification of biased amino acid substitution patterns in human immunodeficiency virus type 1 isolates from patients treated with protease inhibitors. *J. Virol.* **73**:6197–6202.
25. **Shafer, R. W., D. Stevenson, and B. Chan.** 1999. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* **27**:348–352.
26. **Shafer, R. W., M. A. Winters, S. Palmer, and T. C. Merigan.** 1998. Multiple concurrent reverse transcriptase and protease mutations and multidrug resistance of HIV-1 isolates from heavily treated patients. *Ann. Intern. Med.* **128**:906–911.
27. **Spinelli, S., Q. Z. Liu, P. M. Alzari, P. H. Hirel, and R. J. Poljak.** 1991. The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. *Biochimie* **73**:1391–1396.
28. **Sugiura, W., Z. Matsuda, Y. Yokomaku, K. Hertogs, B. Larder, T. Oishi, A. Okano, T. Shiino, M. Tatsumi, M. Matsuda, H. Abumi, N. Takata, S. Shirahata, K. Yamada, H. Yoshikura, and Y. Nagai.** 2002. Interference between D30N and L90M in selection and development of protease inhibitor-resistant human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.* **46**:708–715.
29. **Wang, W., and P. A. Kollman.** 2001. Computational study of protein specificity: the molecular basis of HIV-1 protease drug resistance. *Proc. Natl. Acad. Sci. USA* **98**:14937–14942.
30. **Wlodawer, A., and J. W. Erickson.** 1993. Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Biochem.* **62**:543–585.
31. **Yoshimura, K., R. Kato, M. F. Kavlick, A. Nguyen, V. Maroun, K. Maeda, K. A. Hussain, A. K. Ghosh, S. V. Gulnik, J. W. Erickson, and H. Mitsuya.** 2002. A potent human immunodeficiency virus type 1 protease inhibitor, UIC-94003 (TMC-126), and selection of a novel (A28S) mutation in the protease active site. *J. Virol.* **76**:1349–1358.