A Multivariate Analysis of HIV-1 Protease Inhibitors and Resistance Induced by Mutation

Anna Maria Almerico,* Marco Tutone, Antonino Lauria, Patrizia Diana, Paola Barraja, Alessandra Montalbano, Girolamo Cirrincione, and Gaetano Dattolo

Dipartimento Farmacochimico, Tossicologico e Biologico, Università degli Studi di Palermo, Via Archirafi 32, 90123 Palermo, Italy

Received April 20, 2005

This paper describes the use of the multivariate statistical procedure principal component analysis as a tool to explore the inhibitory activity of classes of protease inhibitors (PIs) against HIV-1 viruses (wild type and more-frequent single mutants, V82A, V82F, and I84V) and against protease enzymes. The analysis of correlations between biological activity and molecular descriptors or similarity indexes allowed a reliable classification of the 51 derivatives considered in this study. The best results were obtained in the case of the I84V mutant for which a high number of predictions was achieved. On this basis, this statistical approach is proposed as a reliable method for the prediction of the activity of PIs, for which the data against mutant strains have not been reported.

1. INTRODUCTION

The treatment regimens for the human immunodeficiency virus type-1 (HIV-1) have included both HIV protease (PR) and reverse transcriptase (RT) enzyme inhibitors, and most of the antiretroviral drugs currently approved for clinical use are directed against one of these targets. The PR is a homodimeric aspartic proteinase, responsible for cleaving precursor proteins that contain the virion structural proteins and enzymes. Highly potent inhibitors of the protease have been developed,[1] and currently, six compounds [ritonavir (RTV), saquinavir (SQV), indinavir (IDV), nelfinavir (NFV), lopinavir (LPV), and amprenavir (APV)] are the only protease inhibitors (PIs) that have received regulatory approval, whereas several other PIs are undergoing clinical trials. These inhibitors bind to the active site of the dimeric PR and were designed to fit the substrate-binding groove of the enzyme. Therefore, they are mainly peptidomimetic compounds and represent transition-state analogues. New generations of PIs have been developed which include nonpeptide inhibitors such as cyclic ureas and other heterocycles. However, although PIs exhibit potent suppression of viral replication in infected patients, their long-term clinical utility is limited by side effects and viral drug resistance. In the case of HIV-1 protease inhibitors, resistance originates from mutations in the protease molecule that lower the affinity of the inhibitors while still maintaining a viable enzymatic profile.[2] Therefore, a high priority for medical research remains the discovery of antiviral agents effective against mutant HIV strains.

As our aim is to develop simple but efficient methods to evaluate, on the basis of chemical−physical descriptors and structural similarity, new compounds that are less likely to trigger resistance or are effective against mutant HIV strains, we recently described the use of multivariate statistical procedures, principal component analysis (PCA) and discriminant analysis (DA), as tools to explore the inhibitory activity of classes of non-nucleoside reverse transcriptase inhibitors against HIV-1 viruses [wild type (wt) and more-frequent mutants, Y181C, V106A, K103N, and L100I] and against RT enzymes.[3] In this paper, we extend this procedure, exploiting the enormous amount of information available on the inhibitory activity of classes of PIs against HIV-1 viruses (wild type and more-frequent single mutants, V82A, V82F, and I84V) and against protease enzymes. PIs have been in the focus of QSAR studies, and many papers have appeared dealing either with the classification of structurally diverse sets of inhibitors with known potencies through computer modeling[4] or with different approaches for generating the set of molecular descriptors.[5,6] However, to date, there have been no attempts to investigate the PIs with respect to resistance induced by mutation by using a multivariate approach.

Biological problems have an intrinsic multivariate nature, involving many variables at the same time, and in general, the relation between these variables and the biological response is hidden and no useful information can easily be extracted. To simplify the data set in a multivariate problem and to obtain an informative picture of the data tendencies, a chemometric multivariate analysis can be used.

In the past, multivariate data analysis has been applied in many fields of science, demonstrating that it is most suitable in handling complex data sets and allowing the investigation of relationships among all objects and all variables simultaneously. In particular, PCA is able to detect similarities among variables and is used to reduce the number of variables, thus, preparing the data for further analysis. The easier mathematical way to represent a multivariate problem is to build a matrix relating variables and objects. In our case, the objects are the PIs, the variables are selected

* Corresponding author tel.: +39 0916161606; fax +39 0916169999; e-mail: almerico@unipa.it.

10.1021/ci050139z CCC: $33.50
Published on Web 11/12/2005

Table 1. Protease K_i (Expressed in nM Concentrations) and Resistance (R)[a]/ susceptibility (S)[b] to Selected Protease Mutants

entry	derivative	log 1/K_i (CLASS)	reference	V82A	V82F	I84V	reference
1	33_DER_CYCLOURETHAN	−0.95 (L)	8a				
2	(R)6H-1,3-DIOS-4D		8b				
3	AG_1350		8c				
4	AHPBA_38		8d				
5	AHPBA_48		8e				
6	AMPRENAVIR	0.77 (H)	8f, 8g	R	S	S	8q, 9a, 9b
7	A_77212		8h				
8	BILA_2185_BS	2.26 (H)	8i				9c
9	BMS_186318	−1.53 (L)	8j				9c
10	CAGEDIM_3C	−3.89 (L)	8k				
11	CYANGUAN_8I	−0.18 (L)	8l				
12	CYCLOSULFAMIDE_25	−0.49 (L)	8m				
13	CYCLOUREA_1	−1.08 (L)	8m				
14	CYCLOUREA_10A	1.57 (H)	8n				
15	CYCPRTR_11	0.33 (L)	8o				
16	DHP_13Y	0.37 (H)	8p				
17	DMP_450	0.47 (H)	8q, 8r	R	R	R	8q, 9d
18	DPC_681	1.92 (H)	8s			S	8s
19	DUPLCTERM_4	0.15 (L)	8t				8t
20	EGCg		8u				
21	ESC_14	−1.30 (L)	8v				
22	HEA_19	2.12 (H)	8w				
23	INDINAVIR	0.28 (L)	8x, 8y	R	R	S	8q, 9b, 9e
24	ISOQUIN_URETH_ANALOG_2		8z				
25	I_12		8aa		R		8aa
26	KNI_272	2.30 (H)	8ab, 8ac				9c, 9f
27	LASINAVIR	0.30 (L)	8ad, 8ae				
28	LC_3		8af				
29	LOPINAVIR	2.72 (H)	8ag, 8s	R	R		9g
30	LY_289612-ANAL1		8ah, 8ai				
31	LY_314163		8aj				
32	LY_326188		8ak				
33	L_689502	1.29 (H)	8al, 8am				
34	MANNAR_9F	−0.52 (L)	8an				8an
35	MANNOPYRANOSIDE_18		8ao				
36	MW_583		8ap				
37	NELFINAVIR	−0.30 (L)	8aq, 8ar	S	S	R	9h, 9c, 9f
38	PALINAVIR	1.51 (H)	8as				9f
39	PARACYCLOPHANE_DER_1	−0.90 (L)	8at				
40	RITONAVIR	0.43 (H)	8g, 8au	S	R	R	8au, 9f, 9c
41	R_87366	−1.04 (L)	8av				
42	SAQUINAVIR	0.60 (H)	, 8aw	S	S	S	8q, 9e
43	SD_146	1.62 (H)	8q, 8ax	S	S	S	8q
44	SE_063	1.52 (H)	8ay				
45	SYN_DIMERIC_DHP_17	−3.68 (L)	8az				
46	TELINAVIR	−0.48 (L)	8ba, 8bb				9c
47	THP_19	1 (H)	8bc				
48	THP_20	1 (H)	8bc				
49	TIPRANAVIR	2.10 (H)	8bd, 8be				9c, 9f
50	TMC_126	−0.04 (L)	8bf				
51	UREATRICL_9	−0.95 (L)	8bg				

[a] R > 2.1-fold resistant ratio (activity on mutant strain/activity on wt). [b] S < 2.1-fold resistant ratio.

chemicals descriptors, and the biological responses are K_i against PR and susceptibility/resistance data. Each object is now placed in an *n*-dimensional space (where *n* is the number of variables). However, it is more practical for us to work in two or three dimensions. The PCA method permits the projection of higher-order space in two or three dimensions with a minimal loss of statistic information. The PCA consists of a rotation process of the original data defined by an **X** matrix of $n \times p$ dimensions, carried out so that the first new axis is oriented in the direction of maximum variance of the data, the second is perpendicular to the first one and is in the direction of successive maximum variance of the data, and so on. Another interesting aspect in a chemometric multivariate analysis is the possibility of classification as done in the DA. The derived classification rule describes a surface which separates the classes, and it may be used to predict class membership.

2. MATERIALS AND METHODS

This study was performed on a combined set of 51 protease inhibitors, both peptidomimetic and nonpeptide compounds, as listed in the Supporting Information (SI-F).

The selection includes most of the derivatives currently present in the database of the National Institute of Allergy and Infectious Diseases (NIAID)[7] and the literature derivatives for which the inhibitory activity against HIV-1 has been reported.[8] Only the most active (representative) compounds belonging to each chemical class were considered. A total of 37 compounds had reported log 1/K_i values (K_i values

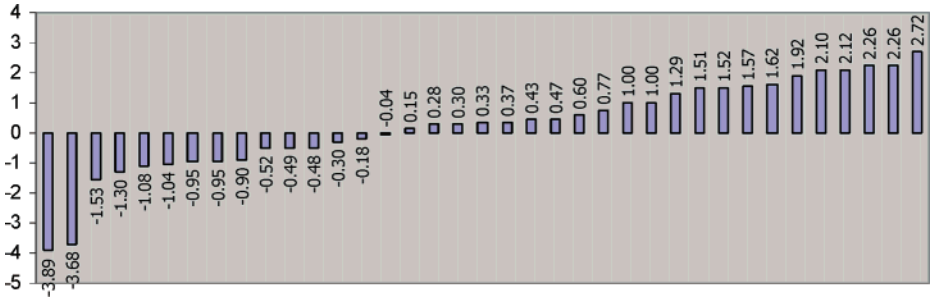

Figure 1. Histogram of the distribution of the activity data for the whole set (cutoff value $1/K_i = 0.35$).

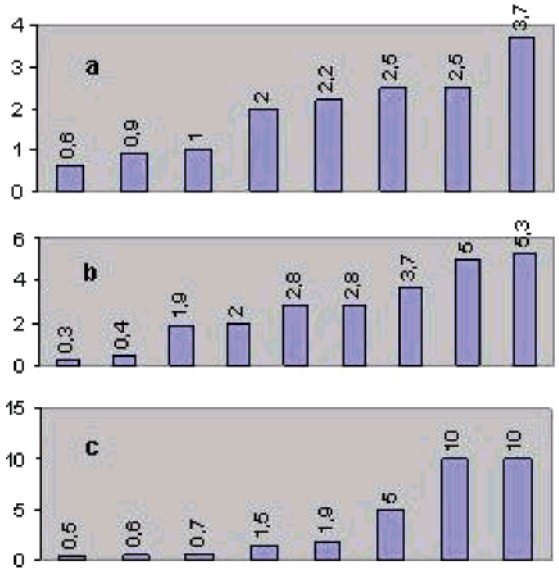

Figure 2. Distribution of PIs' inhibitory activity against selected mutants: (a) V82A, (b) V82F, and (c) I84V.

are in nM), and 14 compounds had no activity data reported (Table 1).

It was decided to classify the HIV-1 PIs on the basis of their enzymatic inhibition. As there were no reported guidelines to use on this data set for the required subdivisions, compounds with log $1/K_i > 0.35$ were considered high active (H) and compounds with log $1/K_i < 0.35$ were considered low active (L). As shown in Figure 1, the log $1/K_i$ values ranged from -3.89 to 2.72, spanning 6 orders of magnitude and with a regular distribution over the whole range.

To classify the same set of compounds on the basis of their activity against the more common PR mutants,[9] derivatives with a resistant ratio (activity on mutant strain/activity on wt) > 2.1 were considered as resistant (R) and compounds with a ratio < 2.1 were considered as susceptible (S). In Figure 2, the distribution of available values against each mutant is shown.

The compounds were sketched as 2D representations using ChemWindow, and the 3D structures were optimized by semiempirical methods (MNDO or PM3 Hamiltonian) using CORINA and COSMIC modules from the TSAR 3.2 software.[10] A pool of more than 80 descriptors was initially calculated for all compounds, from which 16 molecular descriptors, identifying the molecular characteristics that can be related to the biological activity, were selected, and are listed in Table 2.

Our selection of structural variables was made keeping in mind the reported X-ray crystallographic structures of HIV-1 protease-inhibitor complexes which supported the already well-acknowledged binding mode of PIs.[11] In particular, ellipsoidal volume as defined by the moments of inertia, accessible surface area, molecular mass, molecular volume, and molecular refractivity give information about steric properties of a molecule; log *P* is a typical QSAR variable, related to the hydrophobic/hydrophilic profile of the inhibitor and, in our case, can be related to this kind of interaction in the active site; total lipole is a measure of the lipophilic distribution, and it is calculated from the summed atomic log *P* values such as the calculation of the dipole moment;[12] the number of hydrogen-bond donors and acceptors gives other information about the ability of the inhibitor to stabilize its interaction with the "binding pocket"; total dipole μ gives information about the electronic features; total energy in vacuo was already included in chemometric studies.[13] We have also chosen a set of topological indexes, such as flexibility index Φ, which is related to the degree of linearity and the presence of cycles or branching;[14] the number of rotable bonds; the Balaban index;[15] the sum of the E state;[16,17] and the Kier−Hall topological index $^1\chi^v$.[14] These last descriptors encoded information about the flexibility, size, branching, and shape of the molecules.

In our first analysis carried out on derivatives with known log $1/K_i$ values, the correlation matrix on the pool of 16 descriptors revealed coefficients > 0.9 for mass, molecular volume, molecular refractivity, and accessible surface area; therefore, to avoid redundancy in the data set, it was decided to include only the accessible surface area in further analyses. This step reduced the pool to 13 descriptors. Also, $^1\chi^v$ was eliminated because it was uncorrelated to the dependent variable. The reduced descriptor pool (12 descriptors) for all 37 compounds with known log $1/K_i$ values was screened by performing a PCA. Table 3 reports the matrix of the PCs with their composition in terms of original variables, together with the fraction of variance explained, the total fraction of variance explained, and the eigenvalue of the covariance matrix corresponding to each component that is equal to the fraction of variance explained by the number of variables used. The data were standardized by mean/standard deviation. The first four PCs, with eigenvalues >1 and explaining 79.4% of the variance, were selected for further calculations.

Table 3 lists the contribution of molecular descriptors to particular PCs, as also evidenced in Figure 3, which shows the projection of the variables on the factor plane. In PC1, the variables that have major importance are the accessible

Table 2. Molecular Descriptors for All PI Derivatives Reported in Table 1[a]

	descriptors															
entry	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	7.76	560.81	543.72	445.51	2334.60	4.00	5.44	6.16	150.59	15.33	13.00	8	1.10	87.50	2	7
2	−19.12	604.69	538.06	442.45	4882.70	6.22	8.17	11.93	149.67	15.03	9.63	11	1.23	116.48	2	7
3	−4.81	571.86	522.66	442.55	2777.70	7.90	3.19	17.74	160.60	15.85	10.31	10	1.37	85.18	4	5
4	−28.03	649.90	588.82	515.29	3658.70	3.19	6.04	36.39	189.28	16.37	12.69	14	1.39	111.92	4	5
5	−14.37	580.18	465.65	446.04	1682.20	5.73	4.40	20.95	158.80	14.24	9.43	8	1.42	97.06	4	5
6	−18.36	505.69	449.62	368.89	2677.00	8.60	2.66	7.22	134.08	13.33	9.93	12	1.62	85.82	3	7
7	25.38	682.94	624.37	540.80	4785.00	5.47	7.02	8.69	190.24	17.21	15.36	21	1.45	114.67	4	10
8	−10.97	618.91	601.50	483.52	3440.20	6.97	4.42	14.50	177.02	16.27	12.26	13	1.33	95.38	3	6
9	−30.32	670.94	654.52	523.54	5672.60	2.84	3.16	16.49	182.33	16.61	15.43	18	1.50	114.33	5	9
10	114.97	614.84	503.97	477.27	1406.60	1.99	5.36	5.16	178.08	16.74	6.31	10	1.13	96.67	4	6
11	28.03	514.78	466.07	419.96	1683.90	3.61	7.40	2.21	151.76	14.58	9.24	9	1.29	78.50	2	4
12	20.70	634.80	554.47	473.03	2559.70	8.48	4.18	7.08	169.69	16.19	11.11	12	1.26	109.65	4	9
13	25.58	538.69	439.96	420.19	1276.40	6.40	5.69	11.37	152.36	13.47	9.46	10	1.21	89.67	2	5
14	35.88	638.83	527.54	494.78	2196.70	7.46	7.33	5.63	186.89	16.23	9.57	10	0.99	104.33	4	5
15	166.59	730.98	661.40	578.60	5609.20	1.56	4.71	4.69	205.79	18.80	13.08	17	1.06	126.67	6	6
16	−12.44	469.69	436.77	368.95	1231.40	6.12	4.24	2.12	138.28	12.17	7.56	7	1.60	74.47	3	4
17	−10.45	536.73	473.48	422.56	1190.30	6.75	4.94	7.52	159.39	13.59	8.81	8	1.24	90.00	4	3
18	−8.39	669.94	624.30	501.35	5044.50	6.71	4.51	12.81	182.50	17.24	13.91	17	1.65	117.57	5	7
19	34.39	767.02	757.32	605.85	5098.70	1.93	4.93	24.98	213.16	18.76	16.89	19	1.45	135.33	6	8
20	4.88	458.40	381.55	304.55	1360.60	3.40	3.02	3.36	108.95	9.47	6.09	4	1.34	99.83	8	11
21	9.55	574.80	531.55	435.31	3072.10	8.28	2.59	7.01	152.71	15.60	10.50	12	1.38	93.07	2	8
22	−30.20	659.96	634.33	537.20	4103.50	5.41	5.30	21.30	185.80	16.99	14.70	17	1.60	113.08	5	6
23	−17.27	613.88	574.53	484.54	3283.90	4.42	3.46	14.53	175.86	16.01	10.91	12	1.17	99.92	4	7
24	−20.97	605.92	553.13	476.45	2689.40	8.37	3.80	3.73	162.97	17.56	11.03	11	1.39	96.73	3	7
25	−58.66	827.97	667.91	590.81	4336.60	7.34	−1.42	20.15	209.74	19.24	15.74	18	1.25	149.93	7	14
26	−8.13	667.91	576.67	504.83	2848.10	6.42	1.71	19.20	178.54	17.69	13.01	14	1.35	108.68	4	7
27	−22.59	659.91	636.54	517.01	3387.50	4.34	3.34	21.48	178.04	15.80	16.96	20	2.28	112.75	4	9
28	−5.60	702.92	606.24	531.23	4290.80	5.70	6.43	35.81	198.10	17.84	11.65	13	1.11	120.75	3	6
29	−19.59	628.89	584.23	508.33	2282.60	4.02	4.99	19.42	179.36	16.21	13.33	15	1.39	104.50	4	5
30	−52.24	643.86	559.94	491.54	2513.60	9.56	4.02	20.76	177.85	16.92	11.37	13	1.39	113.57	4	6
31	−56.18	708.99	611.36	526.12	2293.60	9.86	2.80	13.97	181.88	20.27	12.71	14	1.46	123.22	4	9
32	−21.58	641.98	547.83	489.14	3624.10	9.05	2.28	8.61	169.37	19.01	11.63	12	1.38	95.00	3	7
33	−21.93	673.93	649.01	531.61	4697.40	5.31	4.77	25.50	188.43	17.33	13.24	16	1.17	110.08	4	8
34	38.26	724.76	648.52	535.96	2536.00	11.41	4.36	14.57	178.28	16.51	11.58	13	1.16	146.67	6	8
35	6.88	297.34	274.55	218.32	476.65	6.34	−0.22	8.65	71.29	6.65	5.40	4	1.86	58.00	4	6
36	3.89	582.75	517.52	448.37	2470.20	7.66	3.67	12.94	162.79	15.01	8.95	11	1.12	98.08	4	6
37	−4.60	567.86	554.21	446.82	3247.50	5.43	4.64	12.62	161.83	15.60	10.60	10	1.38	87.55	4	5
38	−4.95	708.99	650.88	537.17	5730.00	8.71	4.37	23.11	199.65	18.12	13.50	15	1.18	117.58	4	8
39	−29.83	746.01	641.01	577.54	2843.10	2.37	4.39	32.44	203.04	18.46	16.75	12	1.48	128.08	6	8
40	−21.89	721.03	654.06	559.29	4217.90	1.08	5.47	4.03	198.85	18.71	15.76	18	1.35	112.10	4	7
41	−24.57	617.78	526.63	468.58	2286.90	3.20	−0.02	22.42	162.38	14.78	11.09	12	1.38	115.58	5	8
42	−37.39	670.94	599.86	530.33	4227.50	2.94	2.99	16.21	186.17	17.39	12.19	13	1.21	114.58	5	7
43	−3.03	825.01	720.55	623.59	5298.90	10.55	7.60	5.18	233.26	20.47	11.92	12	0.79	141.33	6	7
44	20.33	561.74	474.24	435.37	1469.00	8.13	4.82	8.45	164.70	14.24	8.34	8	1.11	93.17	4	4
45	51.44	773.01	646.76	604.07	3011.00	8.40	6.15	4.93	218.98	19.52	12.44	17	1.17	127.00	2	8
46	−60.07	604.83	555.19	480.58	3836.70	4.86	2.62	19.34	166.65	14.76	12.53	14	1.67	108.42	5	6
47	4.44	531.71	492.98	412.83	2005.00	6.86	5.19	5.23	158.74	13.65	7.64	8	1.10	85.83	3	3
48	36.83	556.72	526.00	422.18	2441.60	5.29	5.07	8.24	164.05	14.29	7.29	8	1.00	89.00	3	4
49	−2.60	602.72	554.90	447.51	5695.80	6.29	6.03	9.64	154.28	15.33	9.67	11	1.23	114.48	2	6
50	2.79	562.74	502.48	407.87	3097.50	9.41	3.47	9.38	144.10	14.79	10.24	13	1.36	93.48	2	9
51	335.51	538.70	473.62	403.02	2805.20	8.25	3.49	12.31	151.34	14.11	6.62	7	0.97	86.92	3	5

[a] Values were obtained by using numerical algorithms included in TSAR 3.2: 1, cosmic total energy (eV); 2, molecular mass (Da); 3, accessible surface area (Å^2); 4, molecular volume (Å^3); 5, ellipsoidal volume (Å^3); 6, total dipole moment (Debye); 7, log *P*; 8, total lipole; 9, molecular refractivity (Å^3); 10, $^1\chi^v$; 11, Φ (shape flexibility); 12, # rotable bonds; 13, Balaban index, 14, Kier electrotopological index; 15, # H donors; and 16, # H acceptors.

surface area (−0.38), Φ (−0.39), and the number of rotable bonds (−0.38). This reflects the importance of the steric approach to the binding pocket and, in particular, the need of a high degree of flexibility to achieve suitable interactions. In PC2, log *P* (+0.48) underlines the importance of lipophilicity in the binding mode due to the presence of hydrophobic pocket S/S′, which allows the correct accommodation of the inhibitor. In the same PC, the Balaban index plays a notable role (−0.54).

PC3 is well-expressed in terms of total dipole μ (−0.75) and the number of hydrogen-bond acceptors (−0.46), together with log *P* (+0.38): this shows that the electronic and lipophilic features and, mainly, the ability to form hydrogen bonds represent a fundamental element for the inhibition process. These considerations can be extended to the fourth PC, where the higher loading coefficients are found for log *P* (+0.48) and the number of hydrogen donors (−0.45). Figure 4 shows a plot of the derivatives against their values for PC1 and PC2.

Since molecular similarity is one of the most useful tools in the computer-aided approach to discover molecules which bind to the same receptor site, we also calculated the shape similarity index. In fact, the shape could provide information on the accessibility of and interaction with the active site.

Meyer showed that the Carbo formula[18] for electron densities can be adapted to measure similarity in terms of molecular shape by means of a simple point-counting algorithm. Each molecule is represented as a system of interlocking spheres with a radius equal to the van der Waals radius, centered on the atomic nuclei. The pair of superimposed molecules is placed in a three-dimensional grid, and a mesh of points is scanned.[19] The following are counted: O_1, the number of grid points lying within molecule 1 only; O_2, the number of grid points lying within molecule 2 only; and B, the number of grid points lying within both molecules. The total number of grid points in each molecule is, thus,

$$T_1 = O_1 + B$$

$$T_2 = O_2 + B$$

The point-counting analogue of the Carbo index is

$$S_{12}{}^{c} = \frac{B}{(T_1 T_2)^{1/2}}$$

Another PCA was performed on the matrix of similarity shape index, and the number of significant PCs (10) was determined as above (eigenvalue > 1; Table 4).

Therefore, the PCs selected for descriptors and the similarity index were separately included in a DA with the aim of providing an assignment to the two classes of activity H or L (see Table 1). By using the Mahalanobis distance discrimination algorithm and the procedure reported by Manly,[20] the TSAR automatic procedure was selected as the stopping procedure. This ends when all the points are correctly classified (no variable will give more than a 5% increase in the total Mahalanobis distance sum between class centers, and the best variable reduces the total number of well-classified points) or all the variables have been added. A data point is considered to be well-classified if the classification rule derived by the DA predicts that point to belong to its true class. A summary of the results achieved with DA is reported in Table 5, which lists the total number of compounds predicted to belong to each class (H or L), broken down by the true class of each individual, and the final set of PCs included in each case, together with the partial hit rate.

The mean of the partial hit rates gives the total hit rate (percentage of statistical units correctly classified by the discriminant function). The latter is an estimate of the true predictive ability of the model, which was validated by Huberty's tests [proportional chance (Cpro) and maximum chance (Cmax)].[21] The proportional chance criterion for assessing model fit is calculated by summing the squared proportion that each group represents of the sample (eq 1). The maximum chance criterion is the proportion of cases in the largest group (eq 2).

$$\mathrm{Cpro} = p^2 + (1 - p)^2 \quad (1)$$

$$\mathrm{Cmax} = (n_L/N_L)100 \quad (2)$$

p is the proportion of subjects in one group, $(1 - p)$ is the proportion of cases in the other group, n_L is the number of subjects in the larger of the two groups, and N_L is the total number of subjects in the combined groups.

Table 3. PCA on the Descriptors Set Considering Only the Entries with Known K_i Values (See Table 1)

Whole Set (37 Derivatives)

descriptor	PC1	PC2	PC3	PC4
1	0.145	0.338	−0.094	−0.515
3	−0.380	0.283	0.057	0.071
5	−0.313	0.238	−0.046	0.114
6	0.126	0.173	−0.751	0.237
7	0.100	0.476	0.377	0.485
8	−0.291	−0.227	−0.024	−0.334
11	−0.393	−0.074	0.137	0.170
12	−0.381	−0.021	0.019	0.188
13	−0.176	−0.547	0.056	0.185
14	−0.344	0.323	−0.134	−0.083
15	−0.284	0.159	0.166	−0.451
16	−0.312	−0.065	−0.459	0.075
fraction of variance explained	0.456	0.160	0.093	0.085
total variance explained	0.456	0.616	0.709	0.794
eigenvalue	5.468	1.922	1.118	1.020

Peptide

descriptor	PC1	PC2	PC3	PC4
1	0.110	−0.445	0.281	−0.288
3	0.423	−0.098	0.020	0.083
5	0.308	−0.269	−0.042	−0.172
6	−0.159	0.285	0.261	−0.254
7	0.013	−0.567	−0.128	0.131
8	0.045	0.088	−0.060	0.748
10	0.351	−0.018	0.222	0.092
11	0.389	0.041	−0.341	0.053
12	0.357	−0.046	−0.311	−0.287
13	−0.017	0.199	−0.667	−0.076
14	0.359	0.199	0.288	0.146
15	0.311	0.174	0.196	0.146
16	0.241	0.443	0.042	−0.311
fraction of variance explained	0.378	0.159	0.128	0.109
total variance explained	0.378	0.537	0.665	0.774
eigenvalue	4.914	2.070	1.661	1.417

Nonpeptide

descriptor	PC1	PC2	PC3
1	−0.086	−0.461	0.269
3	0.444	−0.016	−0.017
5	0.326	−0.094	0.176
6	0.259	−0.204	0.215
7	0.205	0.032	−0.515
8	−0.049	−0.408	0.443
9	0.405	−0.119	−0.203
11	0.333	0.315	0.121
12	0.358	0.230	0.183
13	−0.228	0.497	0.167
15	0.167	−0.350	−0.351
16	0.307	0.185	0.386
fraction of variance explained	0.396	0.177	0.156
total variance explained	0.396	0.573	0.729
eigenvalue	4.751	2.122	1.869

The classification accuracy of the model or minimum hit rate (min. hit rate) should result in a 25% greater value than that achieved by chance (Cpro and Cmax).

The other validation test used in this analysis is Pearson's χ^2 (Chi2).[22] The χ^2 test of statistical significance is a series of mathematical formulas which compare the actual observed frequencies of some phenomenon (in our sample) with the frequencies we would expect if there were no relationship at all between the two variables in the larger (sampled) population. Therefore, χ^2 tests the actual results against the null hypothesis and assesses whether they are different enough to overcome a certain probability that they are due to sampling error.

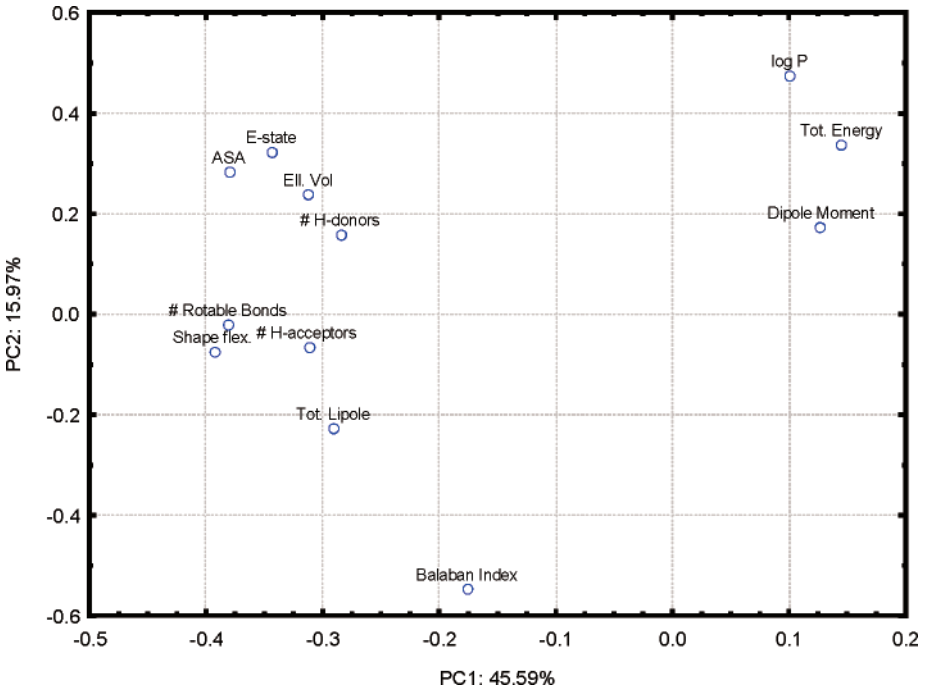

Figure 3. Projection of the variables on the factor plane.

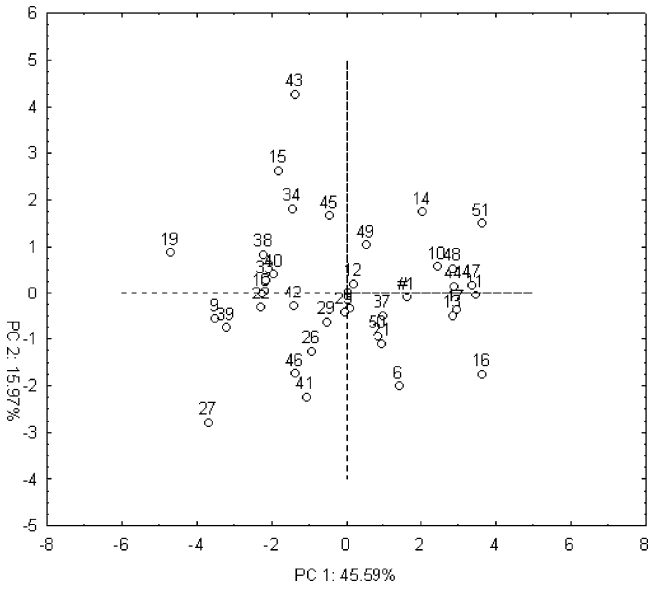

Figure 4. PIs plotted against the first two PCs for descriptors.

3. RESULTS AND DISCUSSION

Data reported in Table 5 show the results of the classification obtained when evaluating log $1/K_i$ of the HIV-1 protease inhibitors. After selecting the appropriate number of PCs as reported in the previous section, the discriminant analyses provided a good assignment of the derivatives to the two classes of activity. In particular, if shape (10 PCs, 81.5% variance explained) is considered, the model correctly classifies both high- and low-activity compounds with a high total hit rate (75.7%), whereas when considering the descriptors (four PCs, 79.4% variance explained), the total hit rate is 73.0%. In any case, on the basis of the requirement for model accuracy (25% better than Cpro and Cmax, 62.6% and 64.1% respectively), our model accuracy rates exceed these standards. The Chi2 test for these models rejected the null hypothesis. The number of entries misclassified is quite similar in both models. Entries **6**, **13**, **26**, **33**, and **45** were always misclassified, probably as a result of their close degree of similarity, whereas entries **11**, **37**, and **42** were misclassified only when the PCs derived from descriptors are used. In all the above-reported cases, as far as descriptors are concerned, the misclassified compounds exhibited an analogous physicochemical profile. Entries **9**, **12**, **16**, and **29** are misclassified only in the case of the shape similarity index as a result of their close degree of similarity.

Even though the obtained results were statistically satisfying, we decided to separately analyze peptidomimetic and nonpeptide compounds, to evidence the difference of the two classes with respect to structure−activity relationships.

Two further data sets were created, the first with the 21 peptidomimetic inhibitors, the second with the 16 nonpeptide inhibitors. Both data sets include only entries with known log $1/K_i$ values, and the same statistic procedures applied previously were again carried out (Tables 3 and 4). For peptidomimetic derivatives, it was possible to obtain a PC matrix using the same set of 12 descriptors as for the whole data set plus ${}^1\chi^v$ (four PCs, 77.4% variance explained), whereas for nonpeptide derivatives, we considered the same 12 descriptors of the whole data set but including molecular refractivity instead of the Kier electrotopological index because the latter was noncorrelated with the dependent variable (three PCs, 72.9% variance explained). This combination of descriptors for both sets of derivatives revealed the best choice to use for further analysis. In an analogous fashion, the similarity shape indexes were considered (peptidomimetics: nine PCs, 78.6% variance explained; nonpeptide: six PCs, 78.3% variance explained).

Peptidomimetic and nonpeptide derivatives were divided into two activity classes on the basis of their enzymatic inhibition using, again, log $1/K_i = 0.35$ as the break point. These data and the selected PCs were analyzed by using the DA. Most relevant results were obtained for nonpeptide

Table 4. PCA on Similarity Shape Index Considering Only the Entries with Known K_i Values (See Table 1)

shape	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
fraction of variance explained	0.238	0.183	0.144	0.060	0.047	0.037	0.031	0.028	0.023	0.022
total variance explained	0.238	0.421	0.566	0.626	0.673	0.710	0.742	0.770	0.793	0.815
eigenvalue	12.147	9.334	7.368	3.069	2.423	1.876	1.603	1.451	1.168	1.108

shape peptide	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
fraction of variance explained	0.238	0.145	0.093	0.072	0.063	0.051	0.046	0.041	0.037
total variance explained	0.238	0.383	0.476	0.549	0.611	0.662	0.708	0.749	0.786
eigenvalue	6.661	4.076	2.593	2.032	1.761	1.428	1.281	1.163	1.036

shape nonpeptide	PC1	PC2	PC3	PC4	PC5	PC6
fraction of variance explained	0.319	0.149	0.118	0.087	0.059	0.050
total variance explained	0.319	0.468	0.587	0.674	0.733	0.783
eigenvalue	7.343	3.432	2.724	1.999	1.356	1.149

Table 5. DA[a] on Activity Classes High and Low

	Descriptors Whole				Descriptors Peptide				Descriptors Nonpeptide			
	predicted				predicted				predicted			
	H	L	total	hit rate	H	L	total	hit rate	H	L	total	hit rate
H	13	5	18	72.2%	7	3	10	70.0%	8	1	9	88.9%
L	5	14	19	73.7%	4	7	11	63.6%	1	6	7	85.7%
total	18	19	37	73.0%	11	10	21	66.7%	9	7	16	87.5%
Cpro[b] = 50.1%; min. hit rate[c] = 62.6%					Cpro[b] = 50.1%; min. hit rate[c] = 62.6%				Cpro[b] = 50.7%; min. hit rate[c] = 63.4%			
Cmax[b] = 51.3%; min. hit rate[c] = 64.1%					Cmax[b] = 52.4%; min. hit rate[c] = 65.4%				Cmax[b] = 56.2%; min. hit rate[c] = 70.2%			
Chi2[d] = 7.79; df = 1; $p < 0.01$					Chi2[d] = 2.37; df = 1; $p < 0.20$				Chi2[d] = 8.90; df = 1; $p < 0.01$			
PC1[e] PC2 PC3 PC4					PC1[e] PC3 PC4				PC3[e]			

	Shape Whole				Shape Peptide				Shape Nonpeptide			
	predicted				predicted				predicted			
	H	L	total	hit rate	H	L	total	hit rate	H	L	total	hit rate
H	13	5	18	72.2%	8	2	10	80.0%	8	1	9	88.8%
L	4	15	19	78.9%	2	9	11	81.8%	2	5	7	71.4%
total	17	20	37	75.7%	10	11	21	80.9%	10	6	16	81.2%
Cpro[b] = 50.1%; min. hit rate[c] = 62.6%					Cpro[b] = 50.1%; min. hit rate[c] = 62.6%				Cpro[b] = 50.7%; min. hit rate[c] = 63.4%			
Cmax[b] = 51.3%; min. hit rate[c] = 64.1%					Cmax[b] = 52.4%; min. hit rate[c] = 65.4%				Cmax[b] = 56.2%; min. hit rate[c] = 70.2%			
Chi2[d] = 9.74; df = 1; $p < 0.01$					Chi2[d] = 8.02; df = 1; $p < 0.01$				Chi2[d] = 6.11; df = 1; $p < 0.025$			

[a] Cross validation employed: leave out one row in turn. [b] Huberty's tests: Cpro (proportional chance), Cmax (maximum chance). [c] min. hit rate: 25% better than chance criteria. [d] χ^2 test; df = degree of freedom; p should be ≤ 0.05 to give statistical significance. [e] Included variables (see Table 3).

derivatives (Table 5), where, for both the descriptors and similarity index (shape), we found remarkable total hit rates (87.5% and 81.2%) that were always higher than the requirements for the criteria of proportional and maximum chance (63.4% and 70.2%). Also in this case, the Chi2 test rejected the null hypothesis for both models. The only derivatives misclassified were **10** and **49** in both models and **21** in the case of shape.

In the case of peptidomimetic compounds (Table 5), the resultant hit rates (80.9% in the case of shape and 66.7% in the case of descriptors) were higher than the requirements of Huberty's tests (Cpro, min. hit rate = 62.6% and Cmax, min. hit rate = 65.4%), but in the case of the Chi2 test, it was not possible to reject the null hypothesis. This remarkable difference between peptidomimetics and nonpeptides in models which consider PCs derived from descriptors can be due to the fact that the set of descriptors used are revealed to be more suitable when the whole set of derivatives or nonpeptide compounds alone are considered.

In the case of nonpeptide derivatives, the major loadings for the PC3 provide evidence that the lipophilicity (log P), joined with the ability to form hydrogen bonds, contributes to determine the main requirements of the binding mode.

However, these data demonstrate that the inhibitory activity of our selected series of PIs can be properly described on the basis of the statistical approach outlined so far.

The next step was to demonstrate that PCA and DA can be used when the resistance or susceptibility against the protease mutant was considered. For this purpose, the available data on the activity against the more common HIV-1 mutants were considered for calculation. These data include the inhibitory activity measured against the cells infected with a mutated virus (EC_{50} values) or against a mutated protease (IC_{50} or K_i values) or other in vitro tests (cf. Table 1).

The chosen mutants were V82A, V82F, and I84V because all of these mutations are in the active site of the enzyme and have been reported in most of the patients receiving treatment with the PIs already approved by the FDA.[1] Moreover, for these mutants, we found, in the literature, the larger portion of data to be available for single mutations: this choice made it simpler to build a model. This time, the PIs were classified as susceptible (S) or resistant (R) against the single mutant strain, as described in Section 2, and the PCA matrices were calculated on the whole data set. The results of the PCA are summarized in Table 6, and significant

Table 6. PCA Matrices Calculated on the Whole Data Set (51 Derivatives; see Table 1)

descriptor	PC1	PC2	PC3	PC4
1	0.142	0.354	0.245	−0.376
3	−0.407	0.249	−0.008	0.050
5	−0.346	0.253	−0.044	0.069
6	0.092	0.054	0.426	0.722
7	0.022	0.562	−0.346	0.043
8	−0.246	-0.153	−0.243	−0.133
11	−0.417	−0.031	−0.194	0.072
12	−0.404	0.049	−0.140	0.137
13	−0.083	−0.538	−0.349	0.109
14	−0.388	0.159	0.269	−0.015
15	−0.224	−0.192	0.381	−0.514
16	−0.285	−0.228	0.426	0.099
fraction of variance explained	0.405	0.165	0.109	0.093
total variance explained	0.405	0.570	0.679	0.772
eigenvalue	4.859	1.979	1.307	1.117

shape	PC1	PC2	PC3	PC4	PC5	PC6
fraction of variance explained	0.209	0.160	0.132	0.058	0.047	0.040
total variance explained	0.209	0.369	0.501	0.559	0.606	0.646
eigenvalue	10.653	8.175	6.730	2.974	2.377	2.058

shape	PC7	PC8	PC9	PC10	PC11	PC12
fraction of variance explained	0.033	0.027	0.026	0.022	0.022	0.020
total variance explained	0.679	0.706	0.733	0.755	0.776	0.796
eigenvalue	1.672	1.387	1.343	1.125	1.110	1.006

Table 7. DA[a] Based on the PCs for Descriptors and Similarity Index for Classes R and S (see Table 6)

	Descriptors				Shape			
	predicted				predicted			
	R	S	total	hit rate	R	S	total	hit rate
				V82A				
R	3	1	4	75.0%	4	0	4	100%
S	1	3	4	75.0%	0	4	4	100%
total	4	4	8	75.0%	4	4	8	100%
	Cpro[b] = 50.0%; min. hit rate[c] = 62.5%				Cpro[b] = 50.0%; min. hit rate[c] = 62.5%			
	Cmax[b] = 50.0%; min. hit rate[c] = 62.5%				Cmax[b] = 50.0%; min. hit rate[c] = 62.5%			
	Chi2[d] = 2; df = 1; $p < 0.20$				Chi2[d] = 8; df = 1; $p < 0.01$			
	PC1[e] PC2							
				V82F				
R	5	0	5	100%	5	0	5	100%
S	2	2	4	50.0%	0	4	4	100%
total	7	2	9	77.8%	5	4	9	100%
	Cpro[b] = 50.6%; min. hit rate[c] = 63.2%				Cpro[b] = 50.6%; min. hit rate[c] = 63.2%			
	Cmax[b] = 55.5%; min. hit rate[c] = 69.4%				Cmax[b] = 55.5%; min. hit rate[c] = 69.4%			
	Chi2[d] = 3.2; df = 1; $p < 0.10$				Chi2[d] = 9; df = 1; $p < 0.01$			
	PC2[e] PC4							
				I84V				
R	3	0	3	100%	3	0	3	100%
S	0	5	5	100%	0	5	5	100%
total	3	5	8	100%	3	5	8	100%
	Cpro[b] = 53.1%; min. hit rate[c] = 66.4%				Cpro[b] = 53.1%; min. hit rate[c] = 66.4%			
	Cmax[b] = 62.5%; min. hit rate[c] = 78.1%				Cmax[b] = 62.5%; min. hit rate[c] = 78.1%			
	Chi2[d] = 8, df = 1, $p < 0.01$				Chi2[d] = 8, df = 1, $p < 0.01$			
	PC2[e] PC3							

[a] Cross validation employed: leave out one row in turn. [b] Huberty's tests: Cpro (proportional chance), Cmax (maximum chance). [c] min. hit rate: 25% better than chance criteria. [d] χ^2 test; df = degree of freedom; p should be ≤ 0.05 to give statistical significance. [e] Included variables (see Table 6).

PCs (four, 77.2% variance explained) were chosen according to the criteria mentioned in Section 2, whereas the DA was carried out considering, in turn, only the derivatives for which the data on resistance and susceptibility were available (cf. Table 1). The results of these discriminant analyses, as reported in Table 7, were generally shown to be very significant statistically since, in four analyses out of six, the hit rate was 100%: all the inhibitors were correctly assigned to the true class. Only in the cases of mutations V82A and V82F, when descriptors were considered, did the Chi2 test not reject the null hypothesis, even though Huberty's test gave statistically significant results. For these mutants, the similarity index seems to give better information for defining the susceptibility or resistance of the derivatives.

In the case of the I84V mutant, PC2 and PC3 support the importance of lipophilicity and hydrogen bonding in the

Table 8. PIs' Classification into Classes R and S[a]

	descriptors			shape		
entry	V82A	V82F	I84V	V82A	V82F	I84V
1	r		r			
2	s	s	r			
3	r			s		
4	r	r	r			
5	r		r			
6	R	S	S	R	S	S
7	s	r	s			
8		s	r			
9	s	r	s			
10		r		r	r	
11	r		r		s	s
12		s	s			
13	r		r			s
14	s			r	s	s
15	s	r	s			
16	r		r	r	r	
17	R	R	R	R	R*	R
18	s	r	S	s	s	S
19	s	r	s	s		
20		r	s			r
21		s		r		
22	s	r	r		s	s
23	R	R	S	R	R	S
24		s				
25		R	s		R	
26		s	s			
27	r	r	r	s		r
28	s	r	r			
29	R	R	r	R	R	r
30		s		s	r	r
31		s	s	s		r
32		s				
33	s	r	s			
34	s	s	s			
35	r		r			
36					r	
37	S*	S	R	S	S	R*
38	s	s	s	s	s	s
39		r	s			
40	S	R	R	S	R	R
41		r	s			
42	S	S*	S	S	S	S
43	S	S	S	S	S	S
44	r			r		r
45	s	s	s			
46	r	r	r	r	r	
47	r		r			
48	r		r			
49	s	s	r			
50		s			s	s
51						

[a] R, S: correct classification. r, s: predicted. R, S: misclassified no-data class. R*, S*: misclassified opposite class.

development of resistance, but the weight of the dipole moment (+0.43) can be related to the relevance of the electronic and spatial nonbonded interactions at the active site. All these data are in good agreement with the structural information available for the nature of resistance.[1]

On the basis of the above results, new calculations were carried out also including, this time, the derivatives for which data against mutants are not available. In the present case, the classification procedure implies the fitting of compounds unclassified against each protease mutant into the classes S or R. Data reported in Table 8 show that it was possible to provide predictions on the activity of derivatives, assigning several compounds to one of the defined classes R or S. The best results were obtained in the case of the I84V mutant, for which the higher number of assignments was found (up to 32 classified molecules).

As far as the compounds' classification is concerned, our results can be summarized as follows.

A comparison of the results for all the protease mutants shows that, in several cases, a unique consistent classification of S was achieved (cf. derivatives **34**, **38**, **45**, **50**) and a classification of R was acheived in the case of compounds **4**, **10**, **16**, **44**, **46**. These findings can be of relevant interest since they can demonstrate that, for the derivatives predicted as S, the presence of any mutations does not necessarily involve an arising of resistance, whereas the opposite becomes truly relevant for the molecules predicted as R. Moreover, in the case of mutation I84V, derivatives **27** and **29** are also always classified as R. These results can be justified considering the high degree of similarity between the molecules which are consistently classified, also confirmed by the analysis of the descriptors which have major importance in the PCs included in the DA. With respect to the derivatives with literature data, they were well classified in the majority of the cases, thus, confirming the relevance of the model. In the case of descriptors, more assignments were possible; this is probably due to the fact that our chosen set takes suitable account of the factors more important in determining resistance or susceptibility.

In the case of mutations V82A and V82F, derivatives with known activity against the selected mutants were well classified with few exceptions. In particular, when considering the V82A mutation, molecules which are classified as R (excluding only derivative **46**) are nonpeptide, and this can suggest that the nonpeptide approach could be of little use for this mutation. In fact, all of the molecules, which are classified as susceptible, are peptidomimetics.

With regard to the approved drugs, it is worthy to note that LPV (**29**) and RTV (**40**) are always correctly classified; APV (**6**) is classified in the opposite class only in the case of I84V.

4. CONCLUSIONS

On the basis of the results reported herein, once again, the multivariate statistical procedure PCA can be proposed as a reliable method for the prediction of the activity of inhibitors, for which the data against mutant strains has not been reported. The approach proposed by us can be used as a sufficiently good and fast discriminator to preliminarily evaluate the probable emergence of resistance to newer synthesized compounds before it is actually verified in biological tests. This in silico screening is not expensive and is easily accessible to most of the researchers active in the field.

Supporting Information Available: Figure reporting the chemical structure of derivatives included in this study (SI-F; PDF format) and Tables with matrix of similarity indexes (ZIP format). This material is available free of charge via the Internet at http://pubs.acs.org

REFERENCES AND NOTES

(1) Swanstrom, R.; Erona, J. Human Immunodeficiency virus type-1 Protease Inhibitors: Therapeutic Successes and Failures, Suppression and Resistance. *Pharmacol. Ther.* **2000**, *86*, 145−170.

(2) Muzammil, S.; Ross, P.; Freire, E. A Major Role for a Set of Non-Active Site Mutations in the Development of HIV-1 Protease Drug Resistance. *Biochemistry* **2003**, *42*, 631−638.

(3) Almerico, A. M.; Lauria, A.; Tutone, M.; Diana, P.; Barraja, P.; Montalbano, A.; Cirrincione, G.; Dattolo, G. A Multivariate Analysis on Nonnucleoside HIV-1 Reverse Transcriptase Inhibitors and Resistance Induced by Mutation. *QSAR Comb. Sci.* **2003**, *22*, 984−996.

(4) See, for example: Patankar, S. J.; Jurs, P. C. Classification of HIV Protease Inhibitors on the Basis of Their Antiviral Potency Using Radial Basis Function Neural Network. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 155−171.

(5) Kiralj, R.; Ferreira, M. M. C. A Priori Molecular Descriptors in QSAR: A Case of HIV-1 Protease Inhibitors: I. The Chemometric Approach. *J. Mol. Graphics Modell.* **2003**, *21*, 435−448.

(6) Katritzky, A. R.; Oliferenko, A.; Lomaka, A.; Karelson, M. Six-membered Cyclic Ureas as HIV-1 Protease Inhibitors: A QSAR Study Based on CODESSA PRO Approach. *Bioorg. Med. Chem. Lett.* **2000**, *12*, 3453−3457.

(7) Web address: http://www.niaid.nih.gov/daids/dtpdb

(8) Data taken from the following papers and reference cited therein: (a) Ghosh, A.; Swanson, L.; Chunfeng, L.; Hussain, K.; Cho, H.; Walters, E.; Holland, L.; Buthod, J. Novel Cyclourethane-Derived HIV Protease Inhibitors: A Ring-closing Olefin Metathesis Based Strategy. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1993−1996. (b) Lee, Y.; Lee, Y.; Lee, J.; Kim, S.; Lee, C.; Park, H. 6-Hydroxy-1,3-dioxin-4-ones as Nonpeptidic HIV Protease Inhibitors. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 2625−2627. (c) Munroe, J.; Hornback, W.; Campbell, J.; Ouellette, M.; Hatch, S.; Muesing, M.; Wiskerchen, M.; Baxter, A.; Su, K.; Campanale, K. LY316340: A Potent HIV-1 Protease Inhibitor Containing a High Affinity Octahydrothienopyridine Hydroxyethylamine Isostere. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 2885−2890. (d) Takashiro, E.; Nakamura, Y.; Miyamoto, S.; Ozawa, Y.; Sugiyama, A.; Fujimoto, K. Design and Synthesis of a Novel Series of HIV-1 Protease Inhibitors. *Bioorg. Med. Chem.* **1999**, *7*, 2105−2114. (e) Takashiro, E.; Hayakawa, I.; Nitta, T.; Kasuya, A.; Miyamoto, S.; Ozawa, Y.; Yagy, R.; Yamamoto, I.; Shibayama, T.; Nakagawa, A.; Yabe, Y. Structure−Activity Relationship of HIV-1 Protease Inhibitors Containing α-Hydroxy-*β*-Amino Acids. Detailed Study of P1 Site. *Bioorg. Med. Chem.* **1999**, *7*, 2063−2072. (f) St. Clair, M.; Millard, J.; Rooney, J.; Tisdale, M.; Parry, N.; Sadler, B.; Blum, M.; Painter, G. In Vitro Antiviral Activity of 141W94 (VX-478) in Combination With Other Antiretroviral Agents. *Antiviral Res.* **1996**, *29*, 53−56. (g) Wilkerson, W.; Akamike, E.; Cheatham, W.; Hollis, A.; Collins, R.; De Lucca, I.; Lam, P.; Ru, Y. HIV Protease Inhibitory Bisbenzamide Cyclic Ureas: A Quantitative Structure−Activity Relationship Analysis. *J. Med. Chem.* **1996**, *39*, 4299−4312. (h) Kempf, D.; Marsh, K.; Paul, D.; Knigge, M.; Norbeck, D.; Kohlbrenner, W.; Codacovi, L.; Vasvanonda, S.; Bryant, P.; Wang, X.; Wideburg, N.; Clement, J.; Plattner, J.; Erickson, J. Antiviral and Pharmacokinetic Properties of C2 Symmetric Inhibitors of the HIV-1 Protease. *Antimicrob. Agents Chemother.* **1991**, *35*, 2209−2214. (i) Doyon, L.; Croteau, G.; Thibeault, D.; Poulin, F.; Pilote, L.; Lamarre, D. Second Locus Involved in Human Immunodeficiency Virus Type 1 Resistance to Protease Inhibitors. *J. Virol.* **1996**, *70*, 3763−3769. (j) Bechtold, C.; Patick, A.; Alam, M.; Greytok, J.; Tino, J.; Chen, P.; Gordon, E.; Ahmad, S.; Barrish, J.; Zahler, R.; Lin, P.; Colonno, R. Antiviral Properties of Aminodiol Inhibitors Against Human Immunodeficiency Virus and Protease. *Antimicrob. Agents Chemother.* **1995**, *39*, 374−379. (k) Hilgeroth, A.; Wiese, M.; Billich, A. Synthesis and Biological Evaluation of the First N−Alkyl Cage Dimeric 4-Aryl-1,4-dihydropyridines as Novel Nonpeptidic HIV-1 Protease Inhibitors. *J. Med. Chem.* **1999**, *42*, 4729−4732. (l) Jhadv, P.; Woerner, F.; Lam, P.; Hodge, N.; Eyermann, C.; Man, H.; Daneker, W.; Bacheler, L.; Rayner, M.; Meek, J.; Erickson-Viitanen, S.; Jackson, D.; Calabrese, J.; Schadt, M.; Chang, C. Nonpeptide Cyclic Cyanoguanidines as HIV-1 Protease Inhibitors: Synthesis, Structure−Activity Relationships, and X-ray Crystal Structure Studies. *J. Med. Chem.* **1998**, *41*, 1446−1455. (m) Schaal, W.; Karlsson, A.; Ahlsen, G.; Lindberg, J.; Andersson, H.; Danielson, H.; Classon, B.; Unge, T.; Samuelsson, B.; Hulten, J.; Hallberg, A.; Karlen, A. Synthesis and Comparative Molecular Field Analysis (CoMFA) of Symmetric and Nonsymmetric Cyclic Sulfamide HIV-1 Protease Inhibitors. *J. Med. Chem.* **2001**, *44*, 155−169. (n) Han, Q.; Chang, C.; Li, R.; Ru, Y.; Jadhav, P.; Lam, P. Cyclic HIV Protease Inhibitors: Design and Synthesis of Orally Bioavailable, Pyrazole P2/P2′ Cyclic Ureas with Improved Potency. *J. Med. Chem.* **1998**, *41*, 2019−2028. (o) Martin, S.; Dorsey, G.; Gane, T.; Hillier, M. Cyclopropane-Derived Peptidomimetics. Design, Synthesis, Evaluation, and Structure of Novel HIV-1 Protease Inhibitors. *J. Med. Chem.* **1998**, *41*, 1581−1597. (p) Hage, S.; Prasad, V.; Boyer, F.; Domagala, J.; Ellsworth, E.; Gajda, C.; Hamilton, H.; Markoski, L.; Steinbaugh, B.; Tait, B.; Lunney, E.; Tummino, P.; Ferguson, D.; Hupe, D.; Nouhan, C.; Grachek, S.; Saunders, J.; VanderRoest, S. Synthesis of 5,6-Dihydro-4-hydroxy-2-pyrones as HIV-1 Protease Inhibitors: The Profound Effect of Polarity on Antiviral Activity. *J. Med. Chem.* **1997**, *40*, 3707−3711. (q) Jadhav, P.; Ala, P.; Woerner, F.; Chang, C.; Garber, S.; Anton, E.; Bacheler, L. Cyclic Urea Amides: HIV-1 Protease Inhibitors with Low Nanomolar Potency against both Wild-Type and Protease Inhibitor Resistant Mutants of HIV. *J. Med. Chem.* **1997**, *40*, 181−191. (r) Hodge, C.; Aldrich, P.; Bacheler, L.; Chang, C.; Eyermann, C.; Garber, S.; Grubb, M.; Jackson, D.; Jadhav, P.; Korant, B.; Lam, P.; Maurin, M.; Meek, J.; Otto, M.; Rayner, M.; Reid, C.; Sharpe, T.; Shum, L.; Winslow, D.; Erickson-Viitanen, S. Improved Cyclic Urea Inhibitors of the HIV-1 Protease: Synthesis, Potency, Resistance Profile, Human Pharmacokinetics and X-ray Crystal Structure of DMP 450. *Chem. Biol.* **1996**, *3*, 301−314. (s) Kaltenbach, R., III; Trainor, G.; Getman, D.; Harris, G.; Garber, S.; Cordova, B.; Bacheler, L.; Jeffrey, S.; Logue, K.; Cawood, P.; Klabe, R.; Diamond, S.; Davies, M.; Saye, J.; Jona, J.; Erickson-Viitanen, S. DPC 681 and DPC 684: Potent, Selective Inhibitors of Human Immunodeficiency Virus Protease Active against Clinically Relevant Mutant Variants. *Antimicrob. Agents Chemother.* **2001**, *45*, 3021−3028. (t) Alterman, M.; Andersson, H.; Garg, N.; Ahlsen, G.; Lovgren, S.; Classon, B.; Danielson, H.; Kvarnstrom, I.; Vrang, L.; Unge, T.; Samuelsson, B.; Hallberg, A. Design and Fast Synthesis of C−Terminal Duplicated Potent C2−Symmetric P1/P1′-Modified HIV-1 Protease Inhibitors. *J. Med. Chem.* **1999**, *42*, 3835−3844. (u) Yamaguchi, K.; Honda, M.; Ikigai, H.; Yukihiko, H.; Shimamura, T. Inhibtory Effects of (−)-Epigallocatechin Gallate on the Life Cycle of Human Immunodeficiency Virus Type 1 (HIV-1). *Antiviral Res.* **1992**, *53*, 19−34. (v) Ghosh, A.; Krishnan, K.; Walters, E.; Cho, W.; Cho, H.; Koo, Y.; Trevino, J.; Holland, L.; Buthod, J. Structure Based Design: Novel Spirocyclic Ethers as Nonpeptidal P2-ligands for HIV Protease Inhibitors. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 979−982. (w) Beaulieu, P.; Wernic, D.; Abraham, A.; Anderson, P.; Bogri, T.; Bousquet, Y.; Croteau, G.; Guse, I.; Lamarre, D.; Liard, F.; Paris, W.; Thibeault, D.; Pav, S.; Tong, L. Potent HIV Protease Inhibitors Containing a Novel (Hydroxyethyl)amide Isostere. *J. Med. Chem.* **1997**, *40*, 2164−2176. (x) Dorsey, B.; Levin, R.; Mcdaniel, S.; Vacca, J.; Guare, J.; Darke, P.; Zugay, J.; Emini, E.; Schleif, W.; Quintero, J.; Lin, J.; Chen, I.; Ostovic, D.; Anderson, P.; Huff, J. L-735,524: The Design of a Potent and Orally Bioavailable HIV Protease Inhibitor. *J. Med. Chem.* **1994**, *37*, 3443−3451. (y) Farese-Di Giorgio, A.; Rouquayrol, M.; Greiner, J.; Aubertin, A.; Vierling, P.; Guedj, R. Synthesis and Anti-HIV Activity of Prodrugs Derived from Saquinavir and Indinavir. *Antiviral Chem. Chemother.* **2000**, *11*, 97−110. (z) Ghosh, A.; Lee, H.; Thompson, W.; Culberson, C.; Holloway, M.; Mckee, S.; Munson, P.; Duong, T.; Smith, A.; Darke, P.; Zugay, J.; Emini, E.; Schlief, W.; Huff, J.; Anderson, P. The Development of Cyclic Sulfolanes as Novel and High-Affinity P2 Ligands for HIV-1 Protease Inhibitors. *J. Med. Chem.* **1994**, *37*, 1177−1188. (aa) Nair, A.; Bonin, I.; Tossi, A.; Welsh, W.; Miertus, S. Computational Studies of the Resistance Patterns of Mutant HIV-1 Aspartic Proteases towards ABT-538 (Ritonavir) and Design of New Derivatives. *J. Mol. Graphics Modell.* **2002**, *21*, 171−179. (ab) Kageyama, S.; Mimoto, T.; Murakawa, Y.; Nomizu, M.; Ford, H., Jr.; Shirasaka, T.; Gulnik, S.; Erickson, J.; Takada, K.; Hayashi, H.; Broder, S.; Kiso, Y.; Mitsuya, H. In Vitro Anti-human Immunodeficiency Virus (HIV) Activities of Transition State Mimetic HIV Protease Inhibitors Containing Allophenylnorstatine. *Antimicrob. Agents Chemother.* **1993**, *37*, 810−817. (ac) Kato, R.; Mimoto, T.; Fukazawa, T.; Morohashi, N.; Kiso, Y. USP 5,932,550. *Chem Abst.* **2001**, *134*, 311436. (ad) Cozens, R.; Bold, G.; Capraio, H.; Fassler, A.; Mestan, J.; Lang, M.; Poncioni, B.; Stover, D.; Rosel, J. Synthesis and Pharmacological Evaluation of CGP 57813 and CGP 61755, HIV-1 Protease Inhibitors from the Phe−C−Phe Peptidomimetic class. *Antiviral Chem. Chemother.* **1996**, *7*, 294−299. (ae) Lazdins, J.; Bold, G.; Capraro, H.; Cozens, R.; Fassler, A.; Flesch, G.; Klimkait, T.; Lang, M.; Mestan, J.; Poncioni, B.; Rosel, J.; Stover, D.; Tintelnot-Blomley, M.; Walker, M.; Woods-Cook, K. Profile of CGP 61755: A Novel and Potent HIV-1 Protease Inhibitor that Shows Enhanced Anti-HIV Activity when Combined with Other Antiretroviral Agents in Vitro. *Schweiz. Med. Wochenschr.* **1996**, *43*, 1849−1851. (af) Marastoni, M.; Bazzaro, M.; Salvatori, S.; Bortolotti, F.; Tomatis, R. HIV-1 P otease Inhibitors Containing an N-Hydroxyamino Acid Core Structure. *Bioorg. Med. Chem.* **2001**, *9*, 939−945. (ag) Molla, A.; Vasavanonda, S.; Kumar, G.; Sham, H.; Johnson, M.; Grabowski, B.; Denissen, J.; Kohlbrenner, W.; Plattner, J.; Leonard, J.; Norbeck, D.; Kempf, D. Human Serum Attenuates the Activity of Protease Inhibitors toward Wild-Type and Mutant Human Immunodeficiency Virus. *Virology* **1998**, *250*, 255−262. (ah) Shepherd, T.; Jungheim, L.; Baxter, A. D-Amino Acids as Novel P2/P3 Ligands for Inhibitors of HIV-1 Protease. *Bioorg. Med. Chem. Lett.* **1994**, *4*, 1391−1396. (ai) Jungheim, L.; Shepherd, T.; Baxter, A.; Burgess, J.; Hatch, S.; Lubbenhusen, P.; Wiskerchen, M.; Muesing, M. Potent Human Immunodeficiency Virus Type 1 Protease Inhibitors That Utilize Noncoded d-Amino Acids as P2/P3 Ligands. *J. Med. Chem.* **1996**,

39, 96−108. (aj) Munroe, J.; Shepherd, T.; Jungheim, L.; Hornback, W.; Hatch, S.; Muesing, M.; Wiskerchen, M.; Su, K.; Campanale, K.; Baxter, A.; Colacino, J. Potent, Orally Bioavailable HIV-1 Protease Inhibitors Containing Noncoded D-Amino Acids. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 2897−2902. (ak) Hornback, W.; Munroe, J.; Shepherd, T.; Hatch, S.; Muesing, M.; Wiskerchen, M.; Colacino, J.; Baxter, A.; Su, K.; Campanale, K. Synthesis and Pharmacokinetics of Potent Carbamate HIV-1 Protease Inhibitors Containing Novel High Affinity Hydroxyethylamine Isosteres. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 2891−2896. (al) Thompson, W.; Fitzgerald, P.; Holloway, M.; Emini, E.; Darke, P.; Mckeever, B.; Schleif, W.; Quintero, J.; Zugay, J.; Tucker, T.; Schwering, J.; Homnick, C.; Nunberg, J.; Springer, J.; Huff, J. Synthesis and Antiviral Activity of a Series of HIV-1 Protease Inhibitors with Functionality Tethered to the P1 or P1′ Phenyl Design. *J. Med. Chem.* **1992**, *35*, 1685−1701. (am) Ho, D.; Toyoshima, T.; Mo, H.; Kempf, D.; Norbeck, D.; Chen, C.; Wideburg, N.; Burt, S.; Erickson, J.; Singh, M. Characterization of Human Immunodeficiency Virus Type 1 Variants with Increased Resistance to a C2-Symmetric Protease Inhibitor. *J. Virol.* **1994**, *68*, 2016−2020. (an) Pyring, D.; Lindberg, J.; Rosenquist, A.; Zuccarello, G.; Kvarnstrom, I.; Zhang, H.; Vrang, L.; Unge, T.; Classon, B.; Hallberg, A.; Samuelsson, B. Design and Synthesis of Potent C2-Symmetric Diol-Based HIV-1 Protease Inhibitors: Effects of Fluoro Substitution. *J. Med. Chem.* **2001**, *44*, 3083−3091. (ao) Murphy, P.; O'Brien, J.; Gorey-Feret, L.; Smith, A., III Structure-based Design and Synthesis of HIV-1 Protease Inhibitors Employing β-d-Mannopyranoside Scaffolds. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1763−1766. (ap) Smith, A., III; Hirschmann, R.; Pasternak, A.; Yao, W.; Sprengeler, P.; Holloway, K.; Kuo, L.; Chen, Z.; Darke, P.; Schleif, W. An Orally Bioavailable Pyrrolinone Inhibitor of HIV-1 Protease: Computational Analysis and X-ray Crystal Structure of the Enzyme Complex. *J. Med. Chem.* **1997**, *40*, 2440−2444. (aq) Kaldor, S.; Kalish, V.; Davies, J.; Shetty, B.; Fritz, J.; Appelt, K.; Burgess, J.; Campanale, K.; Chirgadze, N.; Clawson, D.; Dressman, B.; Hatch, S.; Tatlock, J. Viracept (Nelfinavir Mesylate, AG1343): A Potent, Orally Bioavailable Inhibitor of HIV-1 Protease. *J. Med. Chem.* **1997**, *40*, 3979−3985. (ar) Fujiwara, T.; Sato, A.; El-Farrash, M.; Miki, S.; Abe, K.; Isaka, Y.; Kodama, M.; Wu, Y.; Chen, L.; Harada, H.; Sugimoto, H.; Hatanaka, M.; Hinuma, Y. S-1153 Inhibits Replication of Known Drug-Resistant Strains of Human Immunodeficiency Virus Type 1. *Antimicrob. Agents Chemother.* **1998**, *42*, 1340−1345. (as) Lamarre, D.; Croteau, G.; Wardrop, E.; Bourgon, L.; Thibeault, D.; Clouette, C.; Vaillancourt, M.; Cohen, E.; Pargellis, C.; Yoakim, C.; Anderson, P. Antiviral Properties of Palinavir, a Potent Inhibitor of the Human Immunodeficiency Virus Type 1 Protease. *Antimicrob. Agents Chemother.* **1997**, *41*, 965−971. (at) Ettmayer, P.; Billich, A.; Hecht, P.; Rosenwirth, B.; Gstach, H. Paracyclophanes: A Novel Class of Water-Soluble Inhibitors of HIV Proteinase. *J. Med. Chem.* **1996**, *39*, 3291−3299. (au) Kempf, D.; Marsh, K.; Denissen, J.; Mcdonald, E.; Vasavanonda, S.; Flentge, C.; Green, B.; Fino, L.; Park, C.; Kong, X.; Wideburg, N.; Saldivar, A.; Ruiz, L.; Kati, W.; Sham, H.; Robins, T.; Stewart, K.; Hsu, A.; Plattner, J.; Leonard, J.; Norbeck, D. ABT-538 is a Potent Inhibitor of Human Immunodeficiency Virus Protease and has High Oral Bioavailability in Humans. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2484−2488. (av) Komai, T.; Yagi, R.; Suzuki-Sunagawa, H.; Sakurai, M.; Higashida, S.; Sugano, M.; Handa, H.; Mohri, H.; Yasuoka, A.; Oka, S.; Yabe, Y.; Nishigaki, T.; Kimura, S.; Shimada, K. In Vitro and Ex Vivo Anti-human Immunodeficiency Virus (HIV) Activities of a New Water-soluble HIV Protease Inhibitor, R-87366, Containing (2S,3S)-3-Amino-2-hydroxy-4-phenylbutanoic Acid. *Biol. Pharm. Bull.* **1997**, *20*, 175−180. (aw) Billich, A.; Scholz, D.; Charpiot, B.; Gstach, H.; Lehr, P.; Peichl, P.; Rosenwirth, B. Potent and Orally Bioavailable HIV-1 Proteinase Inhibitors Containing the 2-Aminobenzylstatine Moiety. *Antiviral Chem. Chemother.* **1995**, *6*, 327−336. (ax) Ala, P.; Deloskey, R.; Huston, E.; Jadhav, P.; Lam, P.; Eyermann, C.; Hodge, C.; Schadt, M.; Lewandowski, F.; Weber, P.; Mccabe, D.; Duke, J.; Chang, C. Molecular Recognition of Cyclic Urea HIV-1 Protease Inhibitors. *J. Biol. Chem.* **1998**, *273*, 12325−12331. (ay) Patel, M.; Rodgers, J.; Mchugh, R., Jr.; Johnson, B.; Cordova, B.; Klabe, R.; Bacheler, L.; Erickson-Viitanen, S.; Ko, S. Unsymmetrical Cyclic Ureas as HIV-1 Protease Inhibitors: Novel Biaryl Indazoles as P2/P2′ Substituents. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 3217−3220. (az) Hilgeroth, A.; Billich, A.; Lilie, H. Synthesis and Biological Evaluation of First N-Alkyl Syn Dimeric 4-Aryl-1,4-dihydropyridines as Competitive HIV-1 Protease Inhibitors. *Eur. J. Med. Chem.* **2001**, *36*, 367−374. (ba) Getman, D.; Decrescenzo, G.; Heintz, R.; Reed, K.; Talley, J.; Bryant, M.; Clare, M.; Houseman, K.; Marr, J.; Mueller, R.; Vazquez, M.; Shieh, H.; Stallings, W.; Stegeman, R. Discovery of a Novel Class of Potent HIV-1 Protease Inhibitors Containing the (R)-(Hydroxyethyl)urea isostere. *J. Med. Chem.* **1993**, *36*, 288−291. (bb) Bryant, M.; Getman, D.; Smidt, M.; Marr, J.; Clare, M.; Dillard, R.; Lansky, D.; Decrescenzo, G.; Heintz, R.; Houseman, K.; Reed, K.; Stolzenbach, J.; Talley, J.; Vazquez, M.; Mueller, R. SC-52151, a Novel Inhibitor of the Human Immunodeficiency Virus Protease. *Antimicrob. Agents Chemother.* **1995**, *39*, 2229−2234. (bc) Gayathri, P.; Pande, V.; Sivakumar, R.; Gupta, S. A Quantitative Structure−activity Relationship Study on some HIV-1 Protease Inhibitors Using Molecular Connectivity Index. *Bioorg. Med. Chem.* **2001**, *9*, 3059−3063. (bd) Turner, S.; Strohbach, J.; Tommasi, R.; Aristoff, P.; Johnson, P.; Skulnick, H.; Dolak, L.; Seest, E.; Tomich, P.; Bohanon, M.; Horng, M.; Lynn, J.; Chong, K.; Hinshaw, R.; Watenpaugh, K.; Janakiraman, M.; Thaisrivongs, S. Tipranavir (PNU-140690): A Potent, Orally Bioavailable Nonpeptidic HIV Protease Inhibitor of the 5,6-Dihydro-4-hydroxy-2-pyrone Sulfonamide Class. *J. Med. Chem.* **1998**, *41*, 3467−3476. (be) Molla, A.; Mo, H.; Vasavanonda, S.; Han, L.; Lin, C.; Hsu, A.; Kempf, D. In Vitro Antiviral Interaction of Lopinavir with Other Protease Inhibitors. *Antimicrob. Agents Chemother.* **2002**, *46*, 2249−2253. (bf) Ghosh, A.; Kincaid, J.; Cho, W.; Walters, D.; Krishnan, K.; Hussain, K.; Koo, Y.; Cho, H.; Rudall, C.; Holland, L.; Buthod, J. Potent HIV Protease Inhibitors Incorporating High-affinity P2-ligands and (R)-(Hydroxyethylamino)sulfonamide Isostere. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 687−690. (bg) Han, W.; Pelletier, J.; Hodge, N. Tricyclic Ureas: a New Class of HIV-1 Protease Inhibitors. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 3615−3620.

(9) Data taken from the following papers and from the databases available at the following web addresses: (a) Partaledis, J.; Yamaguchi, K.; Tisdale, M.; Blair, E.; Falcione, C.; Maschera, B.; Myers, R.; Pazhanisamy, S.; Futer, O.; Cullinan, A. In Vitro Selection and Characterization of Human Immunodeficiency Virus Type 1 (HIV-1) Isolates with Reduced Sensitivity to Hydroxyethylamino Sulfonamide Inhibitors of HIV-1 Aspartyl Protease. *J. Virol.* **1995**, *69*, 5228−5235. (b) http://www.hivresistanceweb.com/. (c) http://srdata.nist.gov/hivdb/resistance2_losalamos_HIVPR_inhibitors.html. (d) Klabe, R.; Bacheler, L.; Ala, P.; Erickson-Viitanen, S.; Meck, J. Resistance to HIV Protease Inhibitors: A Comparison of Enzyme Inhibition and Antiviral Potency. *Biochemistry* **1998**, *37*, 8735−8742. (e) http://hivdb.stanford.edu/cgi-bin/PIResiNote.cgi. (f) http://www.mediscover.net/antiviralintro.cfm. (g) Sham, H.; Kempf, D.; Molla, A.; Marsh, K.; Kumar, G.; Chen, C.; Kati, W.; Stewart, K.; Lal, R.; Hsu, A.; Betebenner, D.; Korneyeva, M.; Vasavanonda, S.; McDonald, E.; Saldivar, A.; Wideburg, N.; Chen, X.; Niu, P.; Park, C.; Jayanti, V.; Grabowski, B.; Granneman, G.; Sun, E.; Japour, A.; Leonard, J.; Plattner, J.; Norbeck, D. ABT-378, a Highly Potent Inhibitor of the Human Immunodeficiency Virus Protease. *Antimicrob. Agents Chemother.* **1998**, *42*, 3218−3224. (h) Patick, A.; Mo, H.; Markowitz, M.; Appelt, K.; Wu, B.; Musick, L.; Kalish, V.; Kaldor, S.; Reich, S.; Ho, D.; Webber, S. Antiviral and Resistance Studies of AG1343, an Orally Bioavailable Inhibitor of Human Immunodeficiency Virus Protease. *Antimicrob. Agents Chemother.* **1996**, *40*, 292−297.

(10) All of the calculations were run on a Silicon Graphics Indigo II workstation using the software TSAR 3.2 (Tools for Structure Activity Relationships), VAMP 6.0, and ASP 3.2 (Automated Similarity Packages) (Oxford Molecular-Accelrys). Molecular descriptors were derived according to the method and assumptions reported in the TSAR 3.2 Reference Guide, Oxford Molecular Limited, 1998.

(11) See, for example: (a) Ghosh, A.; Kincaid, J.; Walters, D.; Chen, Y.; Chaudhuri, N.; Thompson, W.; Culberson, C.; Fitzgerald, P.; Lee, H.; McKee, S.; Munson, P.; Duong, T.; Darke, P.; Zugay, J.; Schleif, W.; Axel, M.; Lin, J.; Huff, J. Nonpeptidal P2 Ligands for HIV Protease Inhibitors: Structure-Based Design, Synthesis, and Biological Evaluation. *J. Med. Chem.* **1996**, *39*, 3278−3290. (b) Bardi, J.; Luque, I.; Freire, E. Structure-Based Thermodynamic Analysis of HIV-1 Protease Inhibitors. *Biochemistry* **1997**, *36*, 6588−6596.

(12) Viswanadhan, V.; Ghose, A.; Revankar, G.; Robins, R. Atomic Physicochemical Parameters for Three-Dimensional Structure Directed Quantitative Structure−Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and their Application for an Automated Superposition of Certain Naturally Occurring Nucleoside Antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163−172.

(13) Kubli-Garfias, C. Electronic Structure of Testosterone: A Semiempirical and Ab initio Assessment. *Int. J. Quantum Chem.* **1997**, *62*, 279−289.

(14) Hall, L.; Kier, L. The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure−Property Modelling. In *Reviews in computational chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley: New York, 1992; Chapter 9.

(15) Balaban, A. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, *89*, 399−404.

(16) Hall, L.; Mohney, B.; Kier, L. The Electrotopological State: structure information at the atomic Level for Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 76−82.

(17) Kier, L.; Hall, L. An Electrotopological State Index for Atoms in Molecules. *Pharm. Res.* **1990**, *7*, 801−807.

(18) Carbo, R.; Leyda, L.; Arnau, M. How Similar is a Molecule to another? An Electron Density Measure of Similarity between Two Molecular Structures. *Int. J. Quantum Chem.* **1980**, *17*, 1185−1189.
(19) Meyer, A.; Richards, W. Similarity of a Molecule Shape. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 426−439.
(20) Manly, B. Factor Analysis. In *Multivariate statistical methods. A primer*; Chapman and Hall: New York, 1986: Chapter 7.
(21) Huberty, C. Issues in the Interpretation of Discriminant Analysis. *Psychol. Bull.* **1984**, *95*, 156−169.
(22) Web address: http://www.georgetown.edu/faculty/ballc/webtools/web_chi.html.

CI050139Z