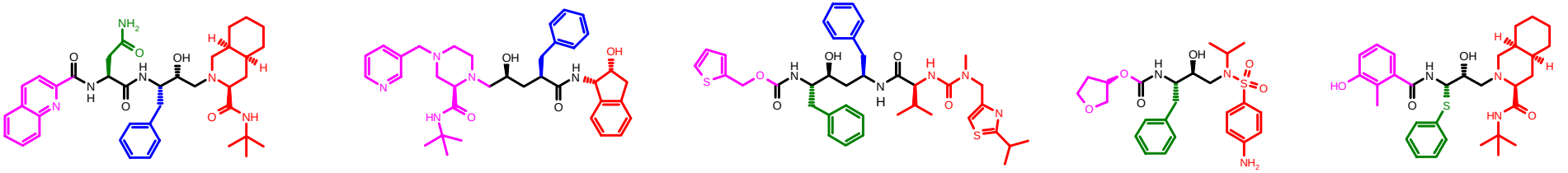
QSAR de alguns inibidores peptídicos da enzima HIV-1 protease utilizando "*a priori*" descritores moleculares e molecular graphics

Rudolf Kiralj (PQ), Márcia Miguel Castro Ferreira (PQ)

rudolf@iqm.unicamp.br, marcia@iqm.unicamp.br

Instituto de Química, Universidade Estadual de Campinas, Campinas 13083-970,SP, Brasil



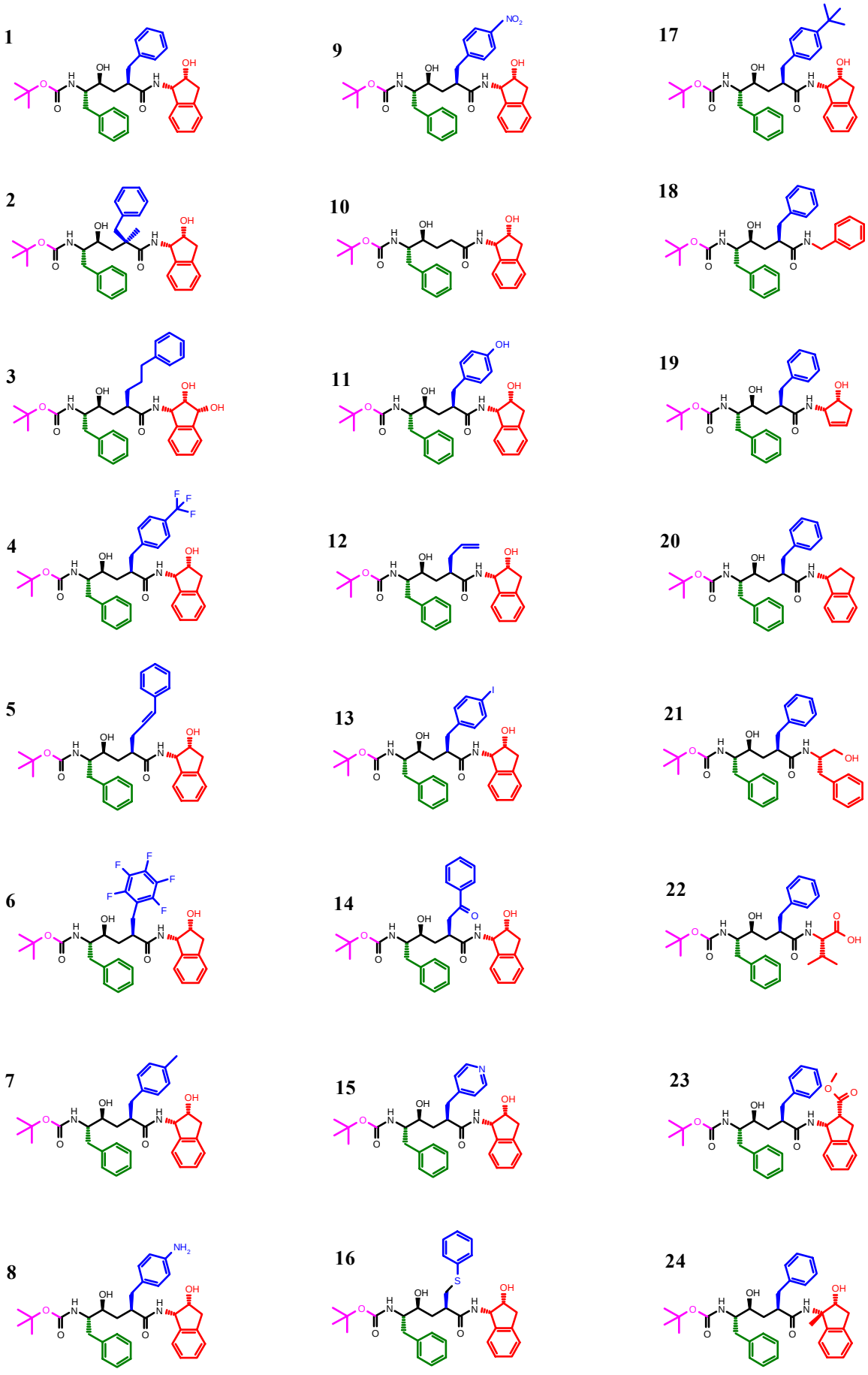


Figure 1a)

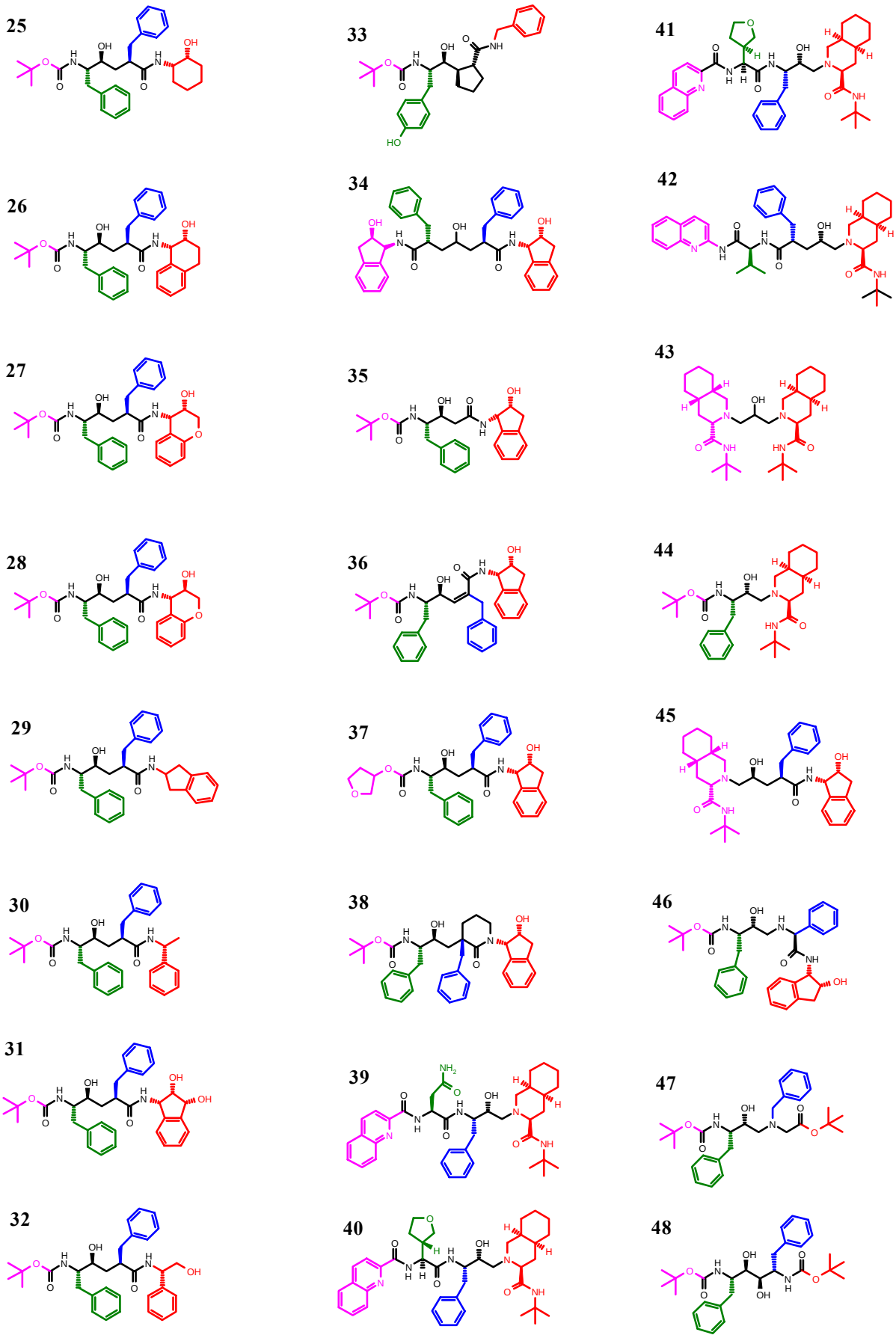


Figure 1 b)

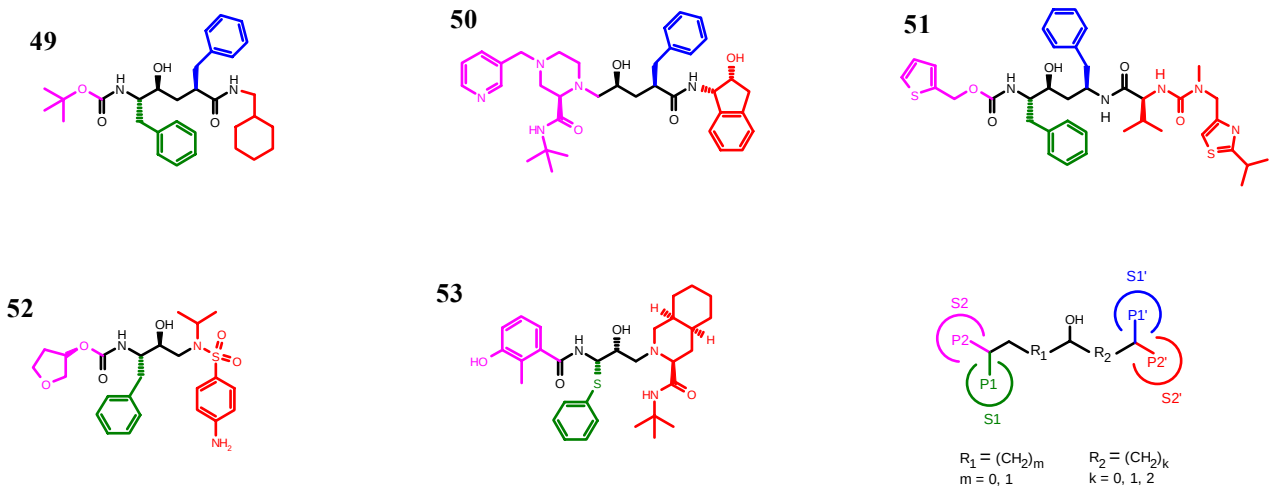


Figure 1 c)

Figure 1. The HIV-1 inhibitors under the study. The substituents P1, P1', P2, P2' coloured differently, in the way treated in this work.

INTRODUCTION

QSAR is an attempt to find a mathematical bridge between a measurable (macroscopic) property of the compounds in biological experiment (biological activity) and their microscopic properties (molecular descriptors).

Questions that arise in this area:

-What molecular descriptors to estimate/calculate, which QSAR methodology to use?

-What softwares to use, sophisticated or simple, free, cheap of expensive?

-How to interpret the result: just showing the quality of the best models, or go into chemistry of the subject, trying to understand the meaning of the results including the meaning of our molecular descriptors?

The main battle is about the black box principle: to accept or not. The second main one is about the quality or quantity to choose: making many QSARs under the black box principle, or performing less QSARs but to understand their full chemistry.

In this work we use *a priori* approach [1], a QSAR methodology where only *a priori* variables ("known before" any sophisticated, computer-assisted calculation) are employed (by hand- or pocket-calculator count/calculation using only 1D and 2D chemical formula). A work on COMBINE (COMparative BINding Energy)-QSAR study on HIV-1 protease inhibitors [2-4] was chosen as a sophisticated QSAR methodology, to demonstrate the reliability and usefulness of our approach on 49 peptide-based hydroxyethylene isostere inhibitors with maximum of four (P_1, P_1', P_2, P_2') substituents (Figure 1) Our results of PCA and HCA (Hierarchical Cluster Analysis) analysis [5,6] and the PLS (Partial Least Squares) prediction [5,6], with the aid of molecular graphics, are discussed in terms both of the *a priori* approach and of the HIV-1 protease inhibitor modeling and are compared to the literature results. The *a priori* approach presented here can be considered as a helpful tool for interpretation of QSARs in terms of basic chemical concepts (molecular size and shape, chemical bonds, atomic properties, electron distribution, hydrogen bonds, effective surface of substituents expressing substituent size, shape, flexibility and polarity responsible for enzyme-substrate interaction, etc.) and as an initial model which can be enriched with various computer-generated descriptors.

METHODOLOGY

STEP 1.
The estimation/calculation&variable selection of the *a priori* molecular descriptors (Tables 1, 2).

STEP 2.
HCA and PCA study of the data.

STEP 3.
PLS prediction of biological activity, Z_1 (total interaction energy) and Z_2 (electrostatic contribution to the free energy of solvation).

STEP 4.
Molecular graphics on the active site of the protease-inhibitor 34 complex.

STEP 1 was based on 2D chemical formula, hand-made chemical schemes and graphs, chemical knowledge and some literature data (cited in Table 1), with a pocket calculator assistance.

STEPs 2 and 3 were performed employing chemometrics softwares Pirouette 3.01 [7] and Matlab 5.4 [8].

STEP 4 utilized molecular graphics softwares Insight II [9] and WebLab Viewer [10] and quantum-chemical MOPAC 6.0 [11] on coordinates of inhibitor 34 [12] and its complex with the HIV-1 protease [13].

RESULTS&DISCUSSION I

The biological activity distribution (Table2):

- the molecules are grouped into three groups: a) 5.158 - 6.246 (molecules 10, 21, 33, 35, 38, 43, 44, 47, 48), 6.640 - 8.268 (molecules 2, 12, 14, 18-20, 22-25, 28-30, 32, 40, 42, 45) a 8.886 - 10.267 (molecules 1, 3-9, 11, 13, 15-17, 26, 27, 31, 34, 36, 37, 39, 41)

- these group can be characterized as slightly active, moderately active, and highly active inhibitors (groups I, II and III, respectively)

Hierarchical Cluster Analysis:

The dendogram on variables (Figure 2a):

- consists of a big cluster(sub-clusters H1-H3) and a small one (H4)

- the two clusters are distinctive due to the internal structure of the data (behavior around *Y vs.* X_i regression line) and the nature of molecular descriptors (H4: molecular size, shape, interactions with no specific direction in space - like hydrophobic; H3: electronic properties like charge distribution and polarity; H2: concepts like molecular size, topology, steric properties, conformational properties; H1: fine details of electronic distribution, especially the role of non-σ electrons involed in aromaticity and heteroaromaticity)

The dendogram on samples (Figure 2b):

-the samples are gropped into two clusters with respect to increase of activity and molecular size: a small cluster is G1 (16 samples: 10, 12, 18-20, 22, 25, 29, 30, 33, 35, 43-45, 47, 48), and a big cluster consisting of three sub-clusters G2 (15 molecules: 1, 2, 5, 7, 14, 17, 21, 23, 24, 26, 28, 32, 36, 38, 46), G3 (9 molecules: 3, 8, 9, 11, 15, 16, 27, 31, 37) and G4 (8 molecules: 4, 6, 13, 34-39, 40, 41, 42)

-the clusters are characterized by distinguished biological activities and structural properties: G1 – small and the smallest and other small molecules with low and moderately high activity; G2 – moderately active molecules; G3 - highly active molecules; G4 – the biggest and other big molecules, mainly highly active

RESULTS&DISCUSSION II

Principal Component Analysis (Fig. 3):

PC's	PC1	PC2	PC3
% Variance	56.49	21.86	7.58
Cum. variance	56.49	78.21	85.79
X_1 or M_r	0.269	0.325	0.234
X_2	0.331	-0.141	-0.086
X_3	0.316	-0.163	-0.260
X_4	0.216	0.405	0.141
X_5	0.224	-0.295	0.244
X_6	0.215	-0.427	-0.163
X_7	0.247	0.352	-0.131
X_8	0.263	0.346	-0.192
X_9	0.292	-0.112	-0.102
X_{10}	0.212	-0.255	0.397
X_{11}	0.285	0.208	-0.130
X_{12} or V_{pol}	0.233	0.016	0.687
X_{13}	0.294	-0.014	-0.188
X_{14}	0.306	-0.224	-0.136

- 3 Principal Components (PC's) enough to describe the inhibitors (86% variance)

- the discriminating role of the PC's: PC1 roughly separates highly active (group III) inhibitors from slightly active ones (group I), while the moderately active are in the middle (group II) as can be observed (Figure 3). The first two PC's confirm the trend found in HCA.

-the chemical background of the PC's: PC1 - meaning biological activity (expressed in terms of molecular size and contents of various types of valence electrons); PC2 - meaning the stereochemical goodness of fit with respect to enzyme (a stereochemical description of the inhibitors); PC3 - meaning the fine (valence electron) distribution of electron density (polar/apolar or hydrophobic/hydrophilic description of the inhibitors)

RESULTS&DISCUSSION III

PLS regression models:

A – Predicting the biological activity:

- PLS results for models I and II (Table 3) use 32 and 48 inhibitors in the training set

-the both models are comparable with those of Pérez *et al.* [14]; the model we propose is a priori model I

- a priori model I is comparable with other literature models:

-the OPTIMOL-MM2X model [2] (r^2=0.78, q^2=0.76, $SDEP_{cv}$=0.68, $SDEP_{ex}$=1.18; our equivalent *a priori* model I including 49 molecule is r^2=0.90, q^2=0.81, $SDEP_{cv}$=0.63, $SDEP_{ex}$=1.68)

- two commercial QSAR softwares of SciVision company: SCIQSAR3.0 [14] (30/8 samples in the training/external validation set, and 5 descriptors in the best model, r^2=0.87, $SDEP_{cv}$=0.50, no other data available) and QSARIS [15] (the best model: 33/15 molecules in the training/validation set, two descriptors (r^2=0.65, q^2=0.57, $SDEP_{cv}$=0.86, $SDEP_{ex}$=1.49), both softwares based on Multiple Linear Regression (MLR)

- a MLR model by Hansch *et al.* [16] (three molecular descriptors, 30 molecules in the training set, r^2=0.82, q^2=0.76, $SDEP_{cv}$=0.69, ratios of regression coefficients and their errors range in 1.3-1.7, other data not available; our equivalent a priori model I is r^2=0.90, q^2=0.80, $SDEP_{cv}$=0.67)

- the prediction of the five clinically approved HIV-1 protease inhibitors 39, 50-53 (Table 4): there are no observed activity data for inhibitors 50-53 measured at the same conditions as for 1-49, and so (the experimental values in Table 3 refer to averaged and normalized data). The predicted values of their activities refer to the group III of highly active inhibitors (with the exception of 52). Underprediction of amprenavir 52 (relatively small inhibitor) by more than one, overprediction of indinavir 50 and ritonavir 51 by 1-2 orders of magnitude in IC_{50} units, can be considered fairly good

RESULTS&DISCUSSION IV

PLS regression models (Tabs. 3, 4):

B – Predicting the energies Z_1 and Z_2:

- Z_1 is well correlated with X_4, X_7-X_9 and X_{11} (48 molecules, 14 variables)

- 3 PC's are enough to describe Z_1, the same as is on biological activity

- PLS model for Z_1, 3 PC's, is quite satisfactory (32/16 molecules in the training/external validation set, 14 variables, q^2=0.76, r^2=0.88, $SDEP_{cv}$=2.21 kcal mol^{-1} across the range of 29.90 kcal mol^{-1})

- Z_2 is correlated with extensive variables X_2, X_3, X_{10} and X_{13} which describe polarity and valence electron distribution,

- PCA with 6 PC's describe Z_2 (over 90% of the variance; 48 molecules, 14 variables)

- PLS model for Z_2 (32 molecules, 14 variables; q^2=0.48, r^2=0.72, $SDEP_{cv}$=0.70 kcal mol^{-1} across a range of 8.84 kcal mol^{-1}) is less quantitative than that for Z_1, but reveals obvious connection between Z_2 and *our a priori* molecular descriptors

Molecular Graphics (Figs. 5, 6):

Figure 5: Crystal structure of HIV-1 protease complexed with inhibitor 34 in various views. The inhibitor Conolly surface is placed inside the electron density isosurface (yellow chicken cage, 0.01 $Å^{-3}$, from PM3-MOPAC 6.0 [11]). The inhibitor indanyl residues lie in the protease pockets S_2, S_2', the phenyl groups are in S_1, S_1'. Many protease residues penetrate the inhibitor isosurface. The molecular space between the Conolly surface and the specified isodensity surface can be considered as the soft (penetrable) molecular volume. The molecular complementarity in the terms of molecular size, shape and functional groups is obvious.

Figure 6: The 29 active site amino-acids (chains A white, B blue) and 10 water molecules around the inhibitor 34 (yellow) at the cut-off distance 5.5 Å (0.1 Å tolerance) with the hydrogen bond (HB) network (green) [1]. The HBs between water, inhibitor and amino-acids contribute to the complex stability: 2 HBs between the catalytic water (left top) and carbonyls of the inhibitor, 2 HBs between this water molecule and two Ile50A, Ile50B, 8 HBs between the inhibitor and the enzyme: 2 between the central OH of the inhibitor and Asp25A, Asp25B; 4 between the OH of indanyl rings of the inhibitor and Asp29A, Asp29B, Gly27A, Gly27B; 2 between the amides of the inhibitor and Gly27A, Gly27B.

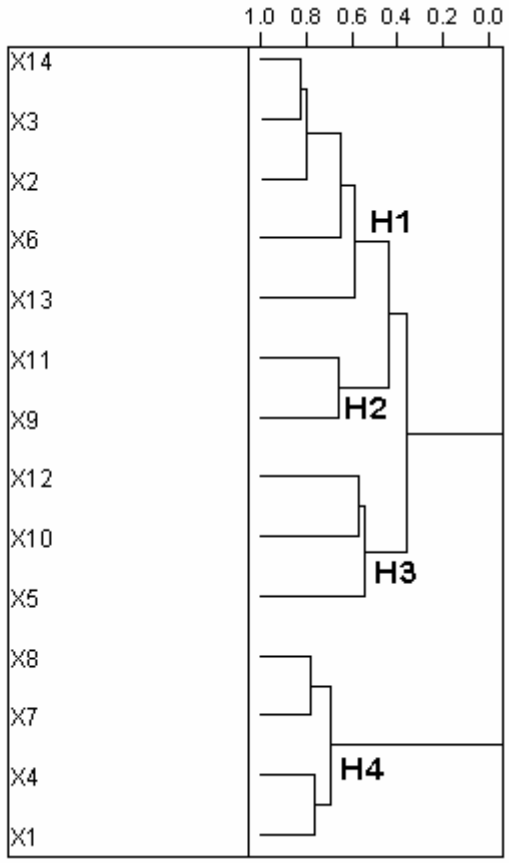


Figure 2a. The HCA dendogram for the *a priori* variables X_1-X_{14}.

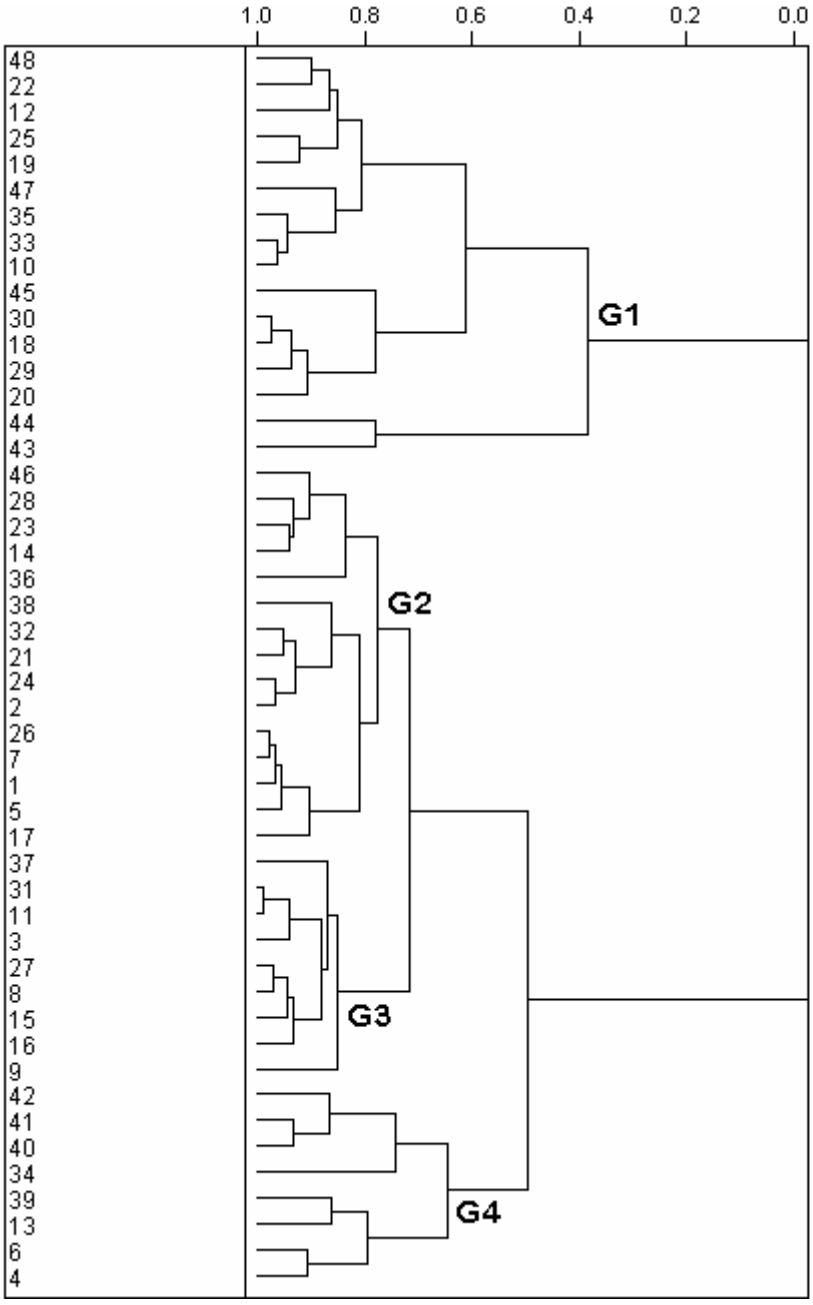


Figure 2b. The HCA dendogram for the samples 1-48.

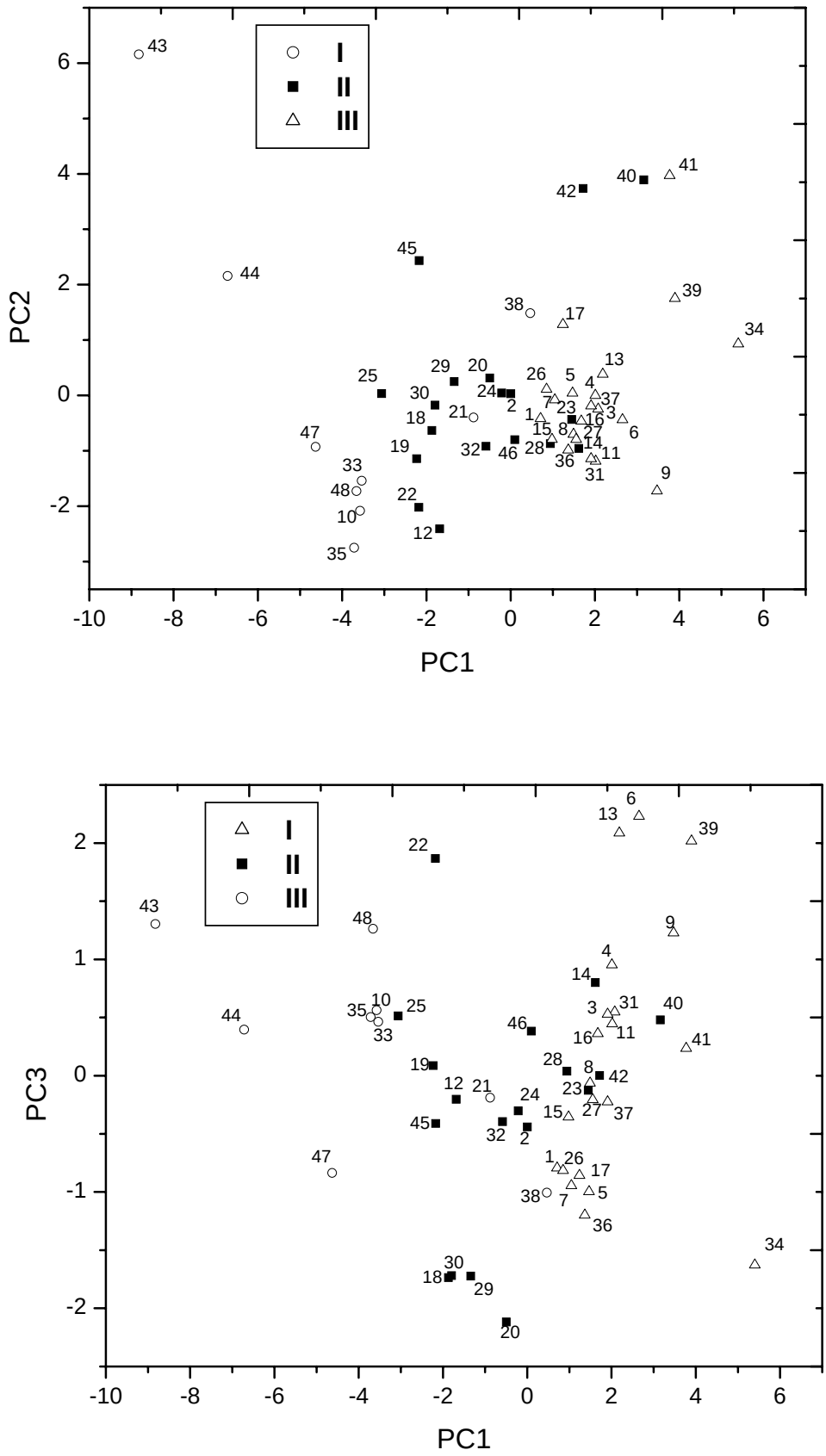


Figure 3. The PCA plots for the samples 1-48, showing the classes I-III.

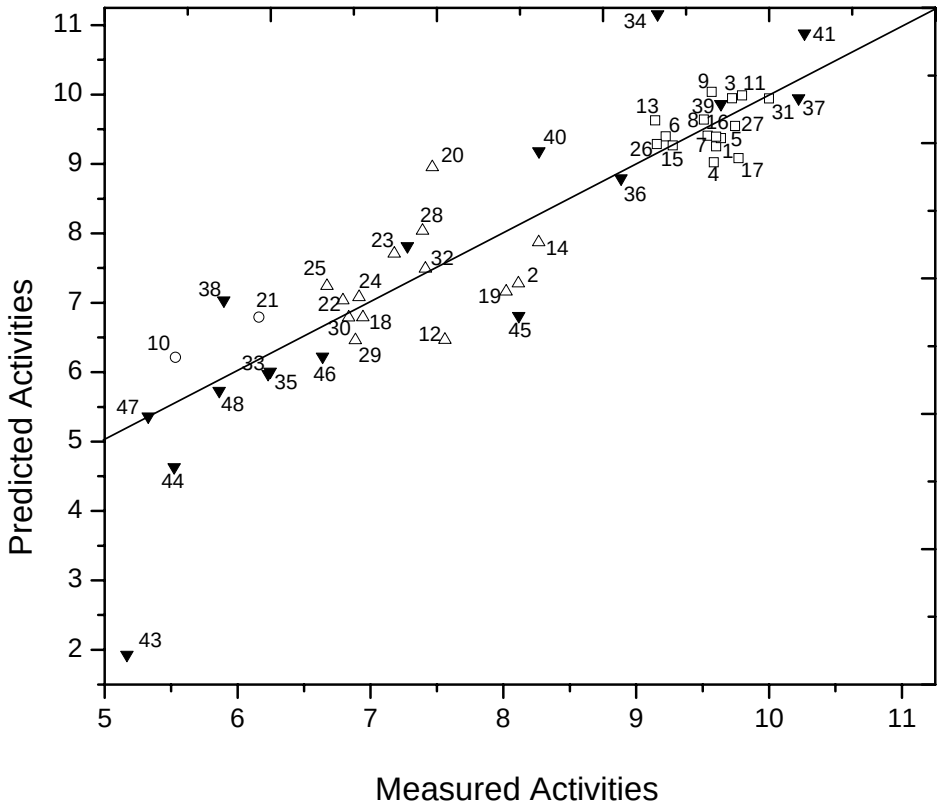


Figure 4. The PLS plot for the *a priori* model I.

CONCLUSION

The biological activity of the peptidic HIV-1 inhibitors under the study:

I - is a three-dimensional phenomena: PC1 - represents biological activity (in terms of molecular size and contents of various types of valence electrons), PC2 - stereochemical fit to enzyme (expressed as molecular branching/compactness and conformation phenomena), PC3 - means fine (valence electron) distribution of electron density (polar/apolar, hydrophobic/hydrophilic relationships inside the inhibitor).

II - is clearly distinguished in three groups of the compounds, as low, moderate and high inhibition activity

III - requires the inhibitors to have all the four substituents aromatic and/or rings

IV - can increase: a)-if both little polar and hydrophobic groups are introduced into the basic structure of the set 1-32, or as alternative, b)-if one or more hooks (flexible hydrophobic chains) are attached on substituents so they enter the active site from the same side of the inhibitor, c)-if more than four (up to 10) substituents are used

The *a priori* molecular descriptors used in this study:

I - are of various chemical nature, like electronic, steric-geometrical, electronic-geometrical, compositional, hydrophobic and topological descriptors

II - well characterized the studied inhibitors and two regression models to predict the activity are comparable with those from literature

III – described also the energetic variables Z1 and Z2, showing that some intrinsic molecular properties are responsible for the behavior of inhibitors in solution

IV - demonstrated how much *a priori* approach can help in chemistry, research and education at low cost

Molecular graphics on inhibitor 34 in this work:

I – illustrated the enzyme-inhibitor molecular complementarity

II – showed that important protease&water-inhibitor interactions occur beyond the classical van der Waals radii

REFERENCES

[1] R. Kiralj, M. M. C. Ferreira, *submitted for publication*.

[2] M. K. Holloway, J. M. Wai, T. A. Halgren, P. M. D. Fitzgerald, J. P. Vacca, , B. D. Dorsey, R. B. Levin, W. J. Thompson, L. J. Chen, S. J. Desolms, N. Gaffin, A. K. Ghosh, E. A. Giuliani, S. L. Graham, J. P. Guare, R. W. Hungate, T. A. Lyle, W. M. Sanders, T. J . Tucker, M. Wiggins, C. M. Wiscount, O. W. Woltersdorf, S. D. Young, P. L. Darke, J. A. Zugay, *J. Med. Chem.*, 38 (1995) 305-317.

[3] M. Pastor, C. Pérez, F. Gago, *J. Mol. Graphics Mod.*, 15 (1997) 363-371.

[4] C. Pérez, M. Pastor, A. R. Ortiz, F. Gago, *J. Med. Chem.*, 41 (1998) 836-852.

[5] M. A. Sharaf, D. L. Illman, B. P. Kowalski: *Chemometrics*, Wiley, New York, 1986.

[6] K. R. Beebe, B. P. Kowalski, *Anal. Chem.*, 59 (1987) 1007A-1017A.

[7] Pirouette 3.01, Infometrix, Inc., Seattle, WA, 2001

[8] Matlab 5.4, The MathWorks, Inc., Natick, MA, 2001.

[9] Insight II, version 95.0. Biosym/MSI, Inc., San Diego, California, 1995.

[10] Web Lab Viewer, version 2.01. Molecular Simulations, Inc., San Diego, California, 1997.

[11] J. J. P. Stewart MOPAC, version 6.0. Frank Seiler Research Laboratory at Air Force Academy, Colorado Springs, Colorado, 1990.

[12] G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, *J. Mol. Biol.*, 267 (1997) 727-748. http://www.ccdc.cam.ac.uk/prods/gold/gold.html

[13] R. Bone, J. P. Vacca, P. S. Anderson, M. K. Holloway, *J. Am. Chem. Soc.*, 113 (1991) 9382-9384. PDB code: 4PHV

[14] SCIQSAR 3.0, SciVision, Burlington, MA, 2000. http://www.scivision.com/gProdPage/tAppNotes/sciQSAR/priori.html

[15] QSARIS, SciVision, Burlington, MA, 2000. http://www.scivision.com/gProdPage/tAppNotes/qsarIS/qsarIS_Note3.html

[16] R. Garg, S. P. Gupta, H. Gao, M. S. Baby, A. K. Debnath, C. Hansch, *Chem. Rev.*, 99 (1999) 3526-3601.

ACKNOWLEDGEMENTS

The authors acknowledge FAPESP for the financial support.

Table 1. Definition and description of the variables.

Symbol	Definition and description
Y	***in vitro*** **inhibition activity, $pIC_{50}= -logIC_{50}$**
X_1 **or** M_r	**relative molecular mass**
X_2	**No. of non-σ valence electrons (the count of π-bonds & the free electrons)**
X_3	**No. of non-hydrogen atoms in planar fragments (in aromatic rings, double bonds)**
X_4	**No. of chemical bonds (excluding hydrogens)**
X_5	**No. of valence electrons per atom**
X_6	**non-σ valence electron surface density X_2/S, S - van der Waals molecular surface area as a sum of literature surface area increments for atoms and groups**
X_7	**No. of non-hydrogen atoms in ring systems (aromatic and aliphatic)**
X_8	**No. of groups CX_n, n=0,1,2,3, X=H or halogen, C from C=O excluded**
X_9	**effective No. of substituents: a) 4 for molecule where the substituents are in position with respect to the central chain line as in 1 (standard molecule); b) if one or two substituents are missing, it is 3 (33, 35, 44-48) or 2 (43), respectively; c) 3.5 if one of the substituents is smaller (12, 18, 19, 22, 25, 30, 32) or in opposite orientation (28, 29, 36) than in the standard; 3.25 (21 and 42) if the substitent is even smaller; d) 3.5 if one of the substituents is sterically hindered by some little group or atom (by CH_3 in 2, 23, 24; by H in 40), or *via* bigger group linked to the main chain (with C=O in 14; with aliphatic ring in 38).**
X_{10}	**No. of potential H-bonds (No. of donors OH, NH, NH_2 + No. of acceptors OH, C=O, -O-)**
X_{11}	**effective No. of ring substituents (aromatic and aliphatic) based on the same rules as for X_9: a) 3 for molecule 1, the standard; b) X_{14}-1 for most of the molecules (1-11, 13-20, 23-33, 35, 36, 38, 39, 44, 46-48) as one substituent is a non-ring system; c) 4 when all the substituents are rings (34, 41); c) 3.5 also for some molecules (37-a small ring substituent, 40-sterically hindered ring); d) 3 also for some molecules (42-a small non-ring substituent, 45-one substituent missing); e) 2.5 also for one molecule (21-a non-ring and a small ring substituent present in the structure); f) 2 also for some molecules (12 and 22-two non-ring substituents present in the structure, 43-only two substituents present and they are rings).**
X_{12} **or** V_{pol}	**van der Waals volume of polar groups (C=O, $-NH_2$, -NH, -N-, $-CF_3$, -S-, -OH, -O-, $-NO_2$, -I) estimated as van der Waals molecular volume as sum of literature volume increments for atoms and groups**
X_{13}	**the length of the total "aromatic vector": No. of atoms in localized, delocalized and aromatic π-systems, and No. of atoms with free electron pairs (N, O, S), and No. of C atoms in CH_m groups (m=1, 2 or 3) which can participate in hyperconjugation all this is summed as L_i for some well defined molecular fragment (L_i=1 if atom is alone); since such fragments are separated with aliphatic groups and are supposed to be independent (orthogonal), they can be understood as aromatic vectors whose summation gives $(\Sigma_i L_i^2)^{1/2}$ and represents the measure of total (hetero)aromaticity**
X_{14}	**similar to X_{13}, the total No. of non-σ electrons that can be involved in "aromatic vectors", including: a) π-electrons of aromatic systems; b) 2 electrons for C=C and C=O bonds; c) 2 electrons for –N- in aliphatic chains; d) 4 electrons for –S-, -O-, -OH; e) eight electrons for $-NO_2$; f) 2 electrons for CH_m (m=1, 2, or 3)**
Z_1	**refined AMBER total interaction energy for HIV-1 protease – inhibitor complexes**
Z_2	**electrostatic contribution to the free energy of solvation of inhibitor**

Table 2. QSAR data for HIV-1 protease inhibitors.

No.	Y	X_1	X_2	X_3	X_4	X_5	X_6/Å^2	X_7/Å^3	X_8/Å^2
1	9.602	544.694	32	30	43	2.650	0.05395	21	31
2	8.113	558.721	32	30	44	2.627	0.05202	21	32
3	9.721	588.748	34	30	46	2.644	0.05287	21	33
4	9.585	612.693	32	31	47	2.843	0.05099	21	32
5	9.638	570.732	33	32	45	2.643	0.05260	21	33
6	9.222	634.647	32	30	48	3.025	0.05174	21	31
7	9.538	558.721	32	31	44	2.627	0.05225	21	32
8	9.509	559.709	33	31	44	2.659	0.05526	21	31
9	9.569	589.692	38	33	46	2.780	0.06140	21	31
10	5.532	454.569	26	23	37	2.657	0.05283	15	24
11	9.796	560.694	34	31	44	2.691	0.05658	21	31
12	7.561	494.634	33	26	38	2.622	0.06074	15	27
13	9.143	670.591	32	30	44	2.725	0.05104	21	31
14	8.266	572.705	35	32	45	2.707	0.05701	21	31
15	9.276	545.682	33	30	43	2.684	0.05640	21	30
16	9.602	576.760	34	30	44	2.691	0.05525	21	31
17	9.770	600.802	32	31	47	2.565	0.04735	21	35
18	6.943	502.657	30	29	39	2.613	0.05309	18	29
19	8.021	494.634	27	26	38	2.622	0.04923	17	27
20	7.465	528.695	30	30	42	2.608	0.05143	21	31
21	6.161	546.710	32	29	42	2.610	0.05203	18	31
22	6.793	512.649	29	26	38	2.623	0.05023	12	26
23	7.179	574.721	35	34	46	2.667	0.05503	21	32
24	6.673	558.721	32	30	44	2.627	0.05202	21	32
25	6.914	510.677	26	22	39	2.557	0.04526	18	28
26	9.155	558.721	32	30	44	2.627	0.05219	22	32
27	9.745	560.694	34	30	44	2.691	0.05663	22	31
28	7.392	560.694	34	30	44	2.691	0.05663	22	31
29	6.886	544.694	30	30	42	2.608	0.05143	21	31
30	6.836	516.684	30	29	40	2.590	0.05116	18	30
31	10.000	560.694	34	30	44	2.691	0.05639	21	31
32	7.413	532.683	32	29	41	2.633	0.05379	18	30
33	6.230	468.596	26	23	36	2.629	0.05076	17	25
34	9.161	618.777	38	38	51	2.705	0.05843	30	37
35	6.246	440.542	26	23	34	2.688	0.05507	15	23
36	8.886	542.679	33	32	43	2.692	0.05638	21	31
37	10.222	558.678	34	30	45	2.734	0.05902	26	31
38	5.897	584.759	32	30	47	2.621	0.05018	27	34
39	9.638	670.856	37	32	53	2.646	0.05037	26	34
40	8.268	683.896	35	28	55	2.602	0.04634	31	37
41	10.267	683.896	35	28	55	2.602	0.04634	31	37
42	7.277	669.912	33	29	53	2.538	0.04398	26	37
43	5.168	532.814	12	8	52	2.319	0.01914	20	29
44	5.523	501.713	19	15	41	2.434	0.03268	16	27
45	8.116	575.795	25	23	38	2.505	0.03915	25	33
46	6.640	559.709	33	30	44	2.659	0.05477	21	31
47	5.328	484.639	26	22	36	2.560	0.04821	12	26
48	5.862	500.638	28	22	37	2.605	0.04949	12	26
49	4.523	508.705	24	22	40	2.494	0.04105	18	30
50	<8.0	613.804	35	30	49	2.609	0.05521	27	34
51	≈8.9	706.943	39	39	53	2.711	0.05273	22	28
52	≈9.2	491.605	30	22	36	2.776	0.05626	17	23
53	≈8.7	538.749	27	22	41	2.476	0.04192	22	29

Table 2. QSAR data for HIV-1 protease inhibitors (continued).

No.	X_9/Å^{-2}	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	Z_1/kcal mol^{-1}	Z_2/kcal mol^{-1}	Y_{pred}
1	4.00	9	3.0	73.4	16.126	48	-80.56	-10.13	9.280
2	3.50	9	2.5	73.4	15.395	46	-76.15	-9.26	7.372
3	4.00	11	3.0	83.8	16.155	52	-84.12	-11.52	9.932
4	4.00	9	3.0	89.0	16.126	48	-82.76	-10.56	9.128
5	4.00	9	3.0	73.4	17.464	50	-82.74	-11.90	9.405
6	4.00	9	3.0	99.4	16.126	48	-79.56	-10.46	9.152
7	4.00	9	3.0	73.4	16.971	50	-81.92	-9.98	9.417
8	4.00	10	3.0	80.9	16.523	50	-81.36	-12.57	9.633
9	4.00	11	3.0	96.9	16.703	56	-84.51	-11.97	9.954
10	3.00	9	2.0	73.4	14.526	40	-67.78	-9.37	5.971
11	4.00	11	3.0	83.8	16.523	52	-81.53	-11.77	9.969
12	3.50	9	2.0	73.4	14.832	44	-74.17	-9.25	6.935
13	4.00	9	3.0	109.1	16.523	48	-83.14	-10.37	9.387
14	3.50	10	2.5	90.9	17.088	50	-81.17	-10.20	7.957
15	4.00	9	3.0	79.6	16.126	48	-81.85	-11.26	9.288
16	4.00	9	3.0	90.9	16.583	52	-80.40	-10.34	9.430
17	4.00	9	3.0	73.4	16.971	50	-85.76	-10.02	9.297
18	3.50	7	2.5	63.0	15.395	44	-73.56	-9.90	6.822
19	3.50	9	2.5	73.4	13.416	44	-75.20	-10.03	7.373
20	4.00	7	3.0	63.0	16.093	44	-77.68	-9.79	8.595
21	3.25	9	2.5	73.4	13.454	46	-70.79	-10.08	6.595
22	3.50	10	2.0	90.9	13.416	42	-69.82	-9.39	6.984
23	3.50	9	2.5	84.2	16.583	54	-75.61	-10.30	7.500
24	3.50	9	2.5	73.4	13.454	46	-78.84	-10.86	7.031
25	3.50	9	2.5	73.4	11.489	40	-74.83	-9.14	7.085
26	4.00	9	3.0	73.4	16.126	48	-81.09	-11.51	9.264
27	4.00	10	3.0	77.1	16.126	52	-82.53	-12.36	9.591
28	3.50	10	2.5	77.1	16.126	52	-76.09	-11.32	7.895
29	3.50	7	2.5	63.0	14.000	42	-76.80	-9.78	6.532
30	3.50	7	2.5	63.0	15.362	42	-75.58	-9.62	6.798
31	4.00	11	3.0	83.8	16.155	52	-82.20	-10.95	9.966
32	3.50	9	2.5	73.4	15.395	46	-74.16	-10.56	7.484
33	3.00	9	2.0	73.4	14.900	38	-65.12	-11.31	6.977
34	4.00	10	4.0	76.6	19.723	60	-88.28	-11.66	11.160
35	3.00	9	2.0	73.4	14.526	40	-61.83	-10.86	6.008
36	3.50	9	2.5	73.4	23.452	48	-79.81	-10.65	8.794
37	4.00	10	3.5	77.1	16.155	52	-83.26	-11.86	9.948
38	3.50	8	2.5	71.1	15.395	46	-66.18	-11.57	7.035
39	4.00	10	3.0	112.6	19.494	52	-86.00	-16.79	9.863
40	3.50	9	3.5	91.3	19.105	50	-81.48	-13.08	9.185
41	4.00	9	4.0	91.3	19.105	50	-91.73	-12.74	10.880
42	3.25	8	3.0	87.6	19.975	42	-80.34	-10.59	7.816
43	2.00	6	2.0	61.6	7.141	20	-73.94	-5.02	1.929
44	3.00	7	2.0	62.4	9.539	28	-70.77	-7.78	4.631
45	3.00	8	3.0	72.6	14.036	38	-80.71	-9.40	6.811
46	3.00	10	2.0	78.6	16.126	50	-72.88	-13.86	6.224
47	3.00	7	2.0	60.9	14.491	38	-68.08	-9.04	5.362
48	3.00	10	2.0	73.6	11.489	40	-66.90	-10.99	5.733
49	3.50	7	2.5	63.0	11.489	34	-72.19	-8.51	6.372
50	4.00	9	3.0	78.2	16.852	50	-	-	-
51	4.00	11	3.5	127.3	20.591	62	-	-	-
52	3.50	10	2.5	82.4	13.675	38	-	-	-
53	3.00	13	3.0	86.6	19.672	38	-	-	-

Table 3. Comparison of *a priori* with literature models.

Model*	samples	variables	PCs	r^2	q^2	$SDEP_{cv}$	$SDEP_{ex}$
C_{amber}	32	48	2	0.89	0.70	0.72	0.83
C_{delphi}	32	47	2	0.90	0.73	0.69	0.59
$C_{expanded}$	48	54	2	0.91	0.81	0.66	-
a priori I	32	14	3	0.91	0.85	0.51	1.12
a priori II	48	14	3	0.87	0.77	0.76	-

*$SDEP_{cv}$ – SDEP (standard error of prediction) of cross-validation, $SDEP_{ex}$ – external SDEP

Table 4. The activities (pIC_{50}) for the five clinically approved inhibitors.

sample	name	Y_{exp}	Y_{pred}
39	saquinavir	9.638	9.863
50	indinavir	8.0	9.370
51	ritonavir	8.9	11.159
52	amprenavir	9.2	7.741
53	nelfinavir	8.7	9.234