ELSEVIER

Journal of Molecular Graphics and Modelling 21 (2003) 499–515

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

A priori molecular descriptors in QSAR: a case of HIV-1 protease inhibitors II. Molecular graphics and modeling☆

Rudolf Kiralj, Márcia M.C. Ferreira*

Instituto de Química, Universidade Estadual de Campinas, Campinas, SP, Brazil

Received 1 May 2002; accepted 14 November 2002

Abstract

Molecular graphics and modeling methods illustrated the chemical background of the a priori approach from part I, and visualized steric and electronic enzyme-inhibitor relationships at qualitative and quantitative level for **34** and its derivatives. The enzyme-inhibitor electron density overlap occurs at 1.5–5.5 Å cut-off distance, beyond van der Waals radii. Derivatives of **34** exhibit linear relationships between biological activity, molecular size and number of intermolecular interactions.

Keywords: A priori molecular descriptors; Quantitative molecular graphics; A priori and quantitative molecular modeling; HIV-1 protease inhibitors; Partial least squares

1. Introduction

Molecular graphics and modeling is today integrated in quantitative structure–activity/property relationship (QSAR/QSPR) studies, protein crystallography and structure-based drug design [1–3]. Molecular modeling, on the other hand, is useful to analyze molecules and molecular systems, to predict molecular and biological properties (QSAR/QSPR), or to contribute to the understanding of molecular interactions in a qualitative and sometimes quantitative way [1]. Visual representations (maps, charts, photographs, pictures) have always been employed to display, characterize and quantify the characteristics of molecules using the same basic geometrical principles of presentation. As X-ray diffraction [4] became a very accurate method employing distances between spots in an X-ray photograph to determine the unit-cell dimensions and atomic interplanar distances in crystals, molecular graphics can be used to measure molecular dimensions and generate descriptors suitable for QSAR [5,6] and QSPR [7,8].

Fourty-eight HIV-1 protease inhibitors were studied by using principal component analysis (PCA), hierarchical cluster analysis (HCA) and partial least squares (PLS) [9] in the companion work to this (part I [10]). Inhibitor L-700,417 (**34** in part I, Fig. 1) was selected to be further analyzed in this work as an illustrative example for: (1) PCA selection of the best views in molecular graphics; (2) empirical measurement of molecular dimensions to calculate molecular volumes and establish electron density–distance relations; (3) counting of environment (enzyme, solvent) atoms around the inhibitor to quantify environment–inhibitor interactions; (4) modeling of inhibitors inside the active site cavity. These methods help to visualize steric, electronic, hydrogen bonding and hydrophobic effects in terms of a priori molecular descriptors (part I), and to evaluate the a priori approach in QSAR. An illustrative example of the QSAR-molecular graphics connection was given by Selassie and Klein [3].

2. Methodology

2.1. Molecular graphics and related methods

2.1.1. Visualization methods

Molecular modeling packages Insight II [11] and WebLab Viewer [12] were employed to make all the color plates (in orthographic projection) of **34** and its complex with HIV-1 protease, both with C_2 symmetry [13]. The coordinates

☆ The paper was presented on the 13th European Symposium on Quantitative Structure–Activity Relationship: Rational Approaches to Drug Design, Düsseldorf, Germany, 27 August–1 September, 2000. The companion paper, Part I, which employs a priori molecular descriptors in QSAR study, appears in volume 21, issue 5, pages 435–448 of this same journal.

* Corresponding author. Tel.: +55-19-3788-3102; fax: +55-19-3788-3023.
E-mail address: marcia@iqm.unicamp.br (M.M.C. Ferreira).

1093-3263/02/$ – see front matter
PII: S1093-3263(02)00202-4

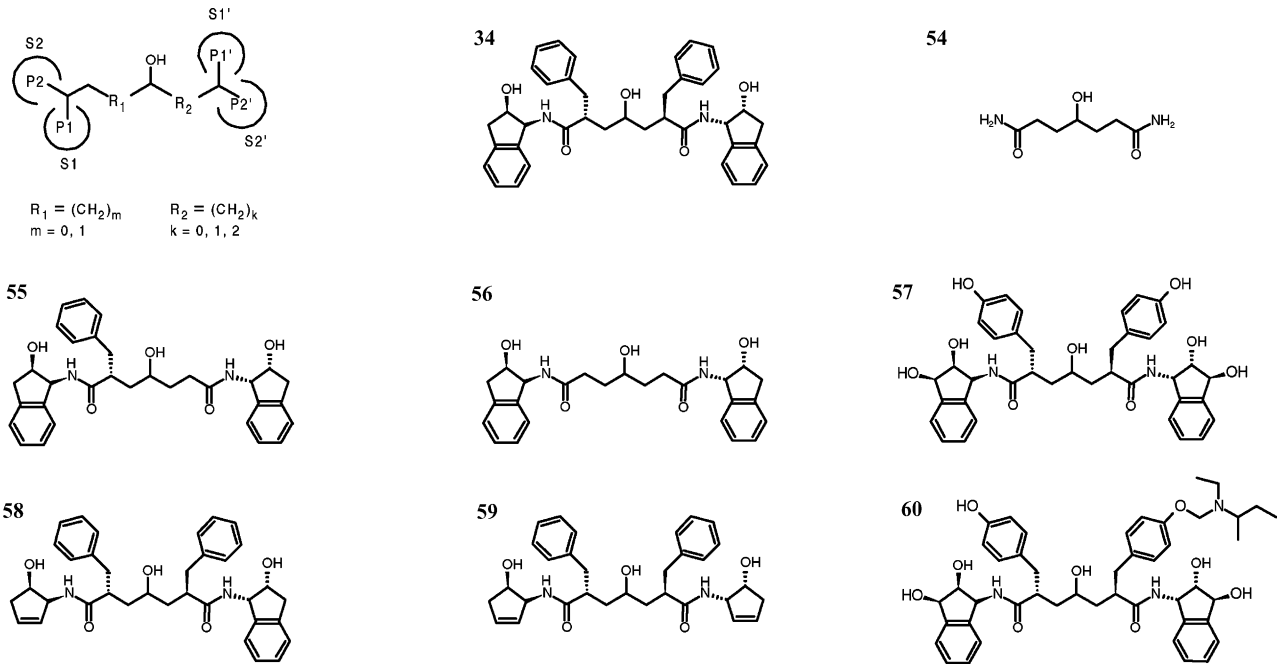


Fig. 1. Two-dimensional representation of HIV-1 protease inhibitors **34**, **54**–**60** and schematic representation of inhibitor side chains (substituents) P_1, P_1', P_2, P_2' and R_1, R_2 separators in the inhibitors. The inhibitors **54**–**60** were modeled by modifying the structure of **34**.

for the free inhibitor [14], free protease [14] and their complex (retrieved from the Protein Data Bank [15]) were used. Species **34** was in disordered orientations 1 and 2 inside the active site [13], and orientation 1 was used for most of purposes. The free inhibitor electron density and HOMO–LUMO isosurfaces (single point calculation using the experimental structure) were generated with MOPAC 6.0 [16,17] using the PM3 Hamiltonian [18,19], and the original inhibitor was manually replaced in the complex by the free inhibitor (displacement 0.3–0.4 Å). A recently developed method [20] for finding the best and the most illustrative molecular views along the axis of principal components (PCA on mean-centered atomic coordinates) or principal moments of inertia (obtained in PM3 polarizability calculation) was used. The method is based on reorientation of initial atomic coordinates with respect to new orthogonal molecular axes which describe most of the variance in atomic coordinates (PCA) or their weighted analogs (weighted by square root of atomic masses in analysis of principal moments of inertia). This way steric (size, shape) and electronic (atom type) effects are pointed out as properties defining molecular orientation.

2.1.2. Empirical measurements of molecular dimensions

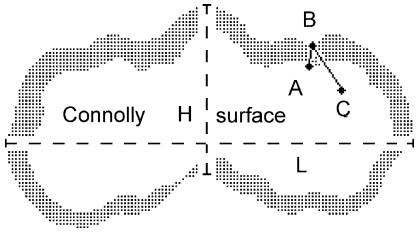
The molecular images from the graphics were projected manually onto a grid paper. A new methodology to calculate volumes V (van der Waals volume, Connolly or surface accessible volume, volumes enclosed in electron isodensity surface at 0.01 and 0.06 e Å^{-3} and in HOMO–LUMO isovalue surface at ± 0.02 Å^{-3}) was performed for **34**. The projected surface area of a molecular image, S_i, was experimentally determined by the grid square method [21] in which the number of unit squares M_i of the projected image (atom, molecule, functional group) onto a 1 mm grid paper were counted manually inside the image boundary. Those at the boundary were considered as 0.5 mm^2. The S_i value was calculated as $S_i = M_i f_i^2$ where f_i is a scale factor determined empirically (based on some well-known molecular dimensions of a fragment in the graphics such as benzene geometry, van der Waals radii, etc.). The measurement (count) of M_i was performed once on each picture if the molecular image was presented in more pictures (in the same orientation and with a different scale factor). In the case where there were unique graphics of an image, the measurement was carried out three times. Various errors can be successfully minimized by using the theoretical area of an internal standard [22]. Thus the measurement errors were estimated as $\varepsilon_i = (S_i/S_0)^{1/2}(f_i/f_0)^2\,\varepsilon_0$ where S_0, f_0, ε_0 refer to the internal standard (van der Waals representation of a side –OH group in color Plate 4, right up, top view), and $\varepsilon_0 = 0.05$ Å^2 is the difference between measured and calculated (using van der Waals radii O: 1.52 Å, H: 1.20 Å from PLATON [23], O–H bond length 0.97 Å [24]) areas. The total measurement error was $\sigma = \text{S.D.} + \langle\varepsilon\rangle$, where S.D. is the standard deviation of S_i, and $\langle\varepsilon\rangle = (\sum_i \varepsilon_i^2)^{1/2}$. Connolly volume of **34** was calculated using the finest grid in Insight II, and the van der Waals volume was estimated employing volume increments [25]. It was then possible to establish a linear V–$\langle S\rangle$ relationship for **34** in the top and bottom view and predict volumes enclosed in the electron isodensity and the HOMO–LUMO isosurfaces presented in the same view.

Since the V–$\langle S\rangle$ relationship was based only on two pieces of data, it was applied to a case of interpolation and another of extrapolation to validate the relationship. The interpolation point was the volume for isodensity at 0.06 $Å^{-3}$, that should be close to the van der Waals volume (according to color plates). The extrapolation case was based on van der Waals volume occupied by 32 overlapped inhibitors from the training set in part I. Two-dimensional projection of the van der Waals volume for the overlapped inhibitors was drawn based on color Plate 1 [26] by Pérez and co-workers [27] and S_i values were measured and averaged. van der Waals volumes of the overlapped inhibitors were also calculated based on volume increments [25] for all the functional groups which appear in the overlap, and also for their different orientations. Both the extrapolation and interpolation points proved the reliability of the V–$\langle S\rangle$ relationship.

To distinguish functional groups such as aromatic, aliphatic, polar H, C, and O atoms at the quantitative level, the 0.01 $Å^{-3}$ electron density isosurface (visible as an envelope) around the Connolly surface (color Plate 4) was graphically divided in such a way that each type of hydrogen, carbon, oxygen from a particular functional group had its own fragment. The lengths (in an orthogonal direction to the Conolly surface) and the projected area of each fragment were measured in all the views. The average thickness Δ for each fragment (see Scheme 1) was determined from the measured data and the thickness data were averaged over the views and functional groups. In some cases, the distance q (see Scheme 1 for the definition) between the point of measurement of Δ (at the electron density isosurface) and the center of the nearest H atom, was measured.

2.1.3. Empirical study of inhibitor–environment intermolecular interactions

The following geometrical parameters of electron density isosurfaces in color Plate 5 (side view, practically the same as in color Plate 4, middle) were measured: projected surface area $\langle S\rangle$ averaged over 10 measurements; the molecular height H and the length L in the same two-dimensional projections (L: the maximum length between two *para*-positioned hydrogens of the P_1, P_1' phenyl groups; H: the maximum height perpendicular to L, passing through the central –OH group, see Scheme 1). Pácios [28] derived the simple Hartree-Fock atomic electron density as a linear combination of exponential functions. Average atomic density of inhibitor **34**, ρ_A, was defined as a simple sum of such atomic densities divided by the number of atoms. The $\log \rho_A$ is shown to be linearly dependent on distance ($r > 0.99$), confirming the expectations that molecular electron density could be modeled using linear molecular dimensions (radii, distances, etc.) regardless of the fact that the atoms are not of the same kind. Linear relationships for $\log \rho$ (ρ-inhibitor electron density) versus linear molecular descriptors ($\langle S\rangle$, $\langle S\rangle^{1/2}$, H, L) had high coefficients of fit ($r > 0.99$). Based upon predicted values for $\rho = 0.01$ and 0.06 e $Å^{-3}$ from the top and bottom view (see sub-Section 2.1.2.), the relationship $V^{1/3}$–$\log \rho$ was used to predict volumes, and $\rho_0 = 0.0653$ e $Å^{-3}$ was estimated for van der Waals volume $V_0 = 553.7$ $Å^3$. Using relationships H–$\log \rho$, and L–$\log \rho$, the dimensions H_0 and L_0 were determined as van der Waals dimensions of the molecule. To define the reference (zero distance) surface, a constant $c = 1.0$ Å was used as a distance from van der Waals surface towards atomic nuclei. In this way, short intermolecular interactions at internuclear distance less than 2 Å involving hydrogens, were included in the study. Since H is measured on hydrogens from polar groups and L on hydrogens from hydrophobic groups, using $d_1 = c + (H - H_0)/2$ and $d_2 = c + (L - L_0)/2$, the mean linear molecular dimension was calculated as $D = \eta_P/(\eta_P + \eta_H)d_1 + \eta_H/(\eta_P + \eta_H)d_2$ where η_P and η_H are numbers of hydrogens from polar (–OH, –NH–) and hydrophobic (–CH_3, –CH_2–,=CH) groups, respectively. Based on these definitions, D has the meaning of "effective" distance from the molecule, which is approximately equal to the distance from atomic nucleus when studying atomic electron density in a particular intermolecular interaction. This approximation is supported by the fact that the relationship between $\log \rho$ and D is obviously linear ($|r| > 0.99$). $\log \rho$–D model for **34** was the basis to study several types of interactions including: (1) whole molecule **34**; (2) only 31 H-bonds by Bone et al. [29]; (3) only hydrophobic groups; (4) only aromatic groups; (5) only benzene carbons; (6) only polar groups. The following intermolecular interaction descriptors were defined: the number of environment valence electrons, N_{ve}, inside cut-off distance D from the inhibitor, counted by using a local routine [30]; average environment electron density, τ, $\tau = N_{ve}/(V - V_0)$ as electron density originated from environment atoms placed inside the soft part of the inhibitor (low electron density); the overlap function, F, between the inhibitor and environment electron density $F = (\rho\tau)^{1/2}(V - V_0)$ averaged over both orientations of the inhibitor. The F–D curve parameters were calculated: the integral I under the curve, the extreme and the mean values. The molecular boundary was predicted based upon these F–D curve parameters.



Scheme 1. Molecular dimensions of inhibitor **34**. The dotted area represents molecular envelope between the Conolly surface and the 0.01 e $Å^{-3}$ electron density isosurface. H and L are molecular height and length, respectively. 'A' and 'B' are the points of measurement of Δ. 'A' is on the Conolly surface (perpendicular to the surface) and 'B' is on the electron density isosurface. 'C' is the center of the nearest hydrogen. The Δ is the distance between 'A' and 'B', and q is the distance between 'B' and 'C'.

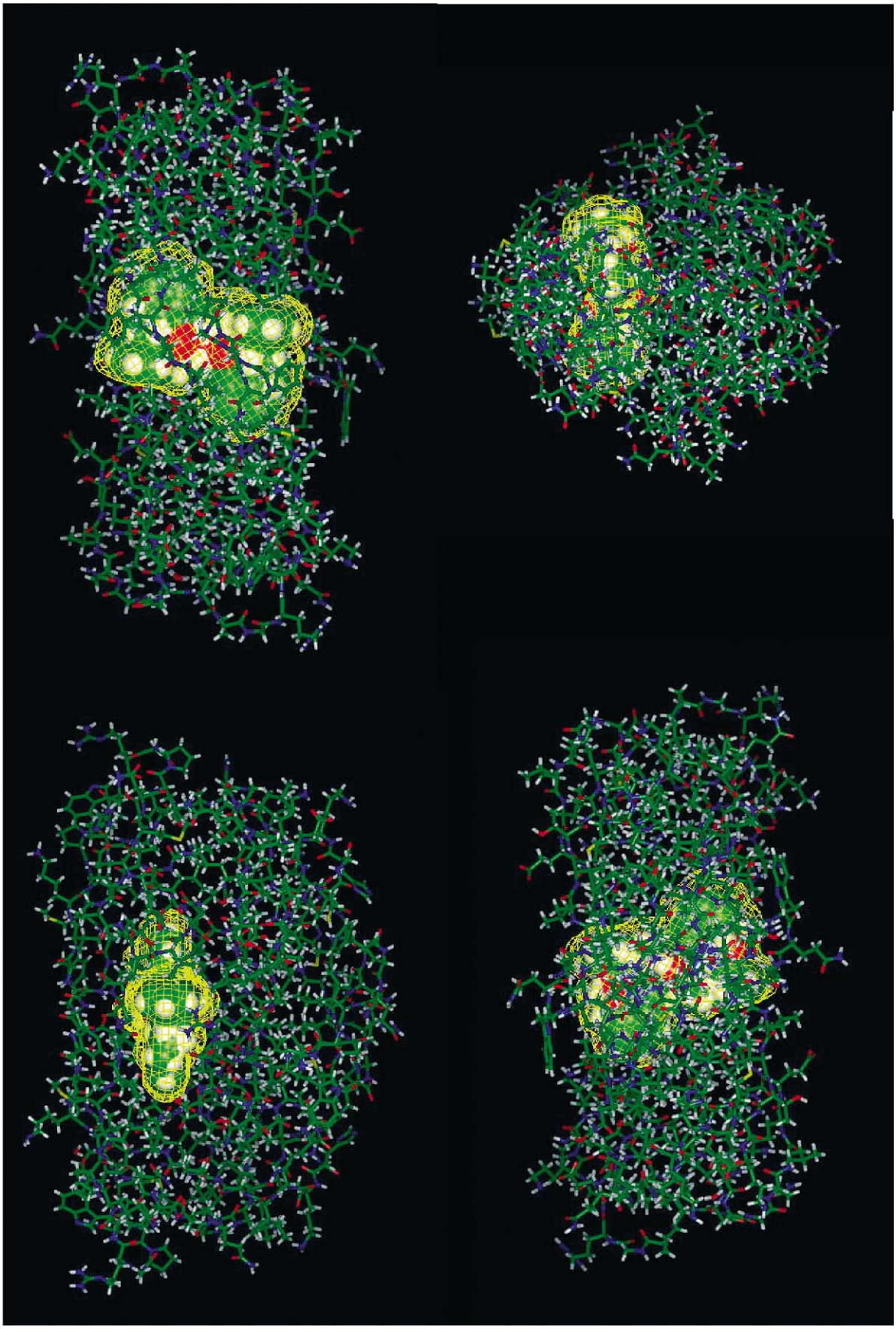


Plate 1. Crystal structure of HIV-1 protease complexed with inhibitor **34** [PDB code: 4PHV] showed as front view (left top), top view (right top), bottom view (left bottom) and side view (right bottom). The approximate C_2 symmetry of the complex can be observed. The atom types are colored in a standard way: carbon—green, hydrogen—white, oxygen—red, nitrogen—blue, sulphur—yellow. The inhibitor Connolly surface is placed inside the electron density isosurface (yellow chicken cage with isovalue 0.01 $Å^{-3}$). Many protease residues penetrate the inhibitor isosurface. The molecular space between the Connolly surface and the specified isodensity surface can be considered as the soft (penetrable) molecular volume.

2.1.3.1. Molecular modeling. Molecular modeling of derivatives from **34** was performed by using WebLab Viewer, modifying the inhibitor (orientation 1) in the structure of the complex. New derivatives **54–59** (Fig. 1) were modeled. Isomers of asymmetrical **55** and **58** (Fig. 1) were named **55a** and **55b**, and **58a** and **58b**, respectively; the isomers **a** and **b** are related to each other *via* a two-fold axis passing through the central OH group. The number N_{HB} of hydrogen bonds was counted. Employing the local routine [30], the number of environment atoms $N_{5.5}$ and its valence electrons N_{ve} (cut-off distance 5.5 Å) were counted. X_9 and X_{10} (see the definition in part I [10]) for **54–59** were calculated. The atom–atom potential parameters [31] were implemented into the local routine [28] and inhibitor–environment interaction energy at 5.5 and 10.0 Å cut-off ($E_{5.5}$ and $E_{10.0}$), the energy per environment atom $\zeta = E_{5.5}/N_{5.5}$, and volume V and area S of **54–60** based on volume and surface area increments [25], were calculated. Approximate biological activity Y_{emp} for **54–60** was empirically estimated (using literature activities [32] for **1–49**, part I) with assumptions: (1) $Y_{emp} = 0$ when there is no substituent (**54**); (2) Y_{emp} is greater or smaller than that of **34** (y_{34}) by some activity increment which depends on the nature of the functional group, its position and number of its chemically equivalent positions. A derivative of inhibitor **57** with flexible hydrophobic chain, an *ortho*-positioned –O–$(CH_2)_2$–N(CH_2CH_3)–CH$(CH_3)CH_2CH_3$ on the P_1 phenyl group (a flexible hook), **60**, was modeled and its empirical activity estimated using various relationships between descriptors for **34**, **54–59**. The log P was calculated by Titan [33] for **34** (PM3 single point calculation using atomic coordinates), **57** and **60** (PM3 single point calculation on coordinates from modeled **57** and **60**). A priori variables (see part I) were calculated for **54–60** and the PLS a priori model I (also from part I) was used to predict the activities for **54–60**. HIV-1 protease-**60** complex was optimized by molecular mechanics MMFF94 [34] in Titan.

3. Results and discussion

The results are presented in Tables 1–5, Fig. 2 and color Plates 1–8.

3.1. Molecular graphics structure considerations

Fig. 2 and color Plates 1–8 illustrate structural properties of HIV-1 protease inhibitor, applied to **34** (L-700,417, color Plate 4).

3.1.1. Visualization methods

The principal axes of inertia defined as directions of the best views in molecular graphics can be utilized. It is known that the principal axes coincide with the symmetry axes [36]. The protease, the protease inhibitor complex and the active site possess an approximate C_2 symmetry axis, and the protease has two approximate mirror planes intersecting in the axis (see Color Plates). In Table 1 and in color Plate 1 three roughly orthogonal views are defined. This orthogonality between the views is illustrated by PCA and PM3 transformations of original atomic coordinates of the inhibitor, enzyme and the complex. Transformations of the free inhibitor atomic coordinates [14] by means of PCA and PM3 polarizability calculation (see Section 2) give very similar orientations which coincide with the top and bottom and the side views, although there is some difference for the third orientation (the front view defined in color Plate 4). The two approaches (principal axes of inertia and PCA) are practically identical. The relative contribution of the three PCs in describing the total variance can aid in better understanding steric and topological properties of the molecular system under study, as follows: (1) the free inhibitor: PCs describe 70.6, 25.0 and 4.4% of the total variance corresponding to the top and bottom, the side and front view; (2) inhibitor from the complex [13]: similarly to (1), the PCs contain 71.3, 24.7, 4.0%. PC3 represents the side view in where all the aromatic rings have uniform thickness; (3) inhibitor-active site amino-acids complex: the complex (color Plate 2) is more spherical than the inhibitor (PCs contain 42.0, 35.6 and 22.4% of the total variance); (4) the complex HIV-1 protease inhibitor shown in four projections (color Plates 1 and 3)—top and bottom (related to each other by 180° rotation), side and front views. PCs describe 61.0, 25.2 and 13.8% of the total variance. The PCs correspond to the front, top and bottom and the side view, respectively. There is a slight difference between the side view and the PC3 direction; (5) free protease: the PCs contain 55.0, 30.5 and 13.7%, demonstrating that the free enzyme form has significantly different conformation than the bound form [14].

3.1.2. Empirical measurements of molecular dimensions

Table 1 contains the results of empirical measurements of areas as two-dimensional projections of van der Waals and Connolly volume or volume enclosed in electron isodensity and frontier orbital isosurfaces. Relatively high errors of measurements (2–18%) are mostly due to manual positioning of the objects, while repeated measurements on the same picture have reasonable smaller errors (<6%). The projected HOMO–LUMO isosurface and 0.01 $Å^{-3}$ electron isodensity surface are of the same size in all views. van der Waals and 0.06 $Å^{-3}$ electron isodensity are practically of the same size in the top and bottom projection. The surface sizes and their ratios are not constant in the three views, and the surfaces follow the increasing order: van der Waals–Connolly–0.01 $Å^{-3}$ electron density ≈0.02 $Å^{-3}$ HOMO–LUMO. The largest differences between the surfaces are related mostly to the side view, revealing that this is the direction which is highly sensitive to changes in volume.

The envelope of electron isodensity at 0.01 $Å^{-3}$ around Connolly volume (color Plate 4) shows significantly different behavior in various views, and depends also on the

Table 1
Empirical measurement of two-diensional projections of van der Waals, Connolly surface accessible, electron isodensity and frontier orbital isodensity volumes for inhibitor **34**

Color plate	View	Type	S_i (Å^2) (ε_i (Å^2))	$\langle S\rangle$ (Å^2) (σ (Å^2))
2 (right up)	Front	van der Waals	94.84 (0.21)	99.41 (14.54)
4 (right down)			85.20 (0.17)	
Not presented[a]			102.54 (1.69)	
7 (left)			115.06 (0.87)	
1 (left up)		Connolly	112.16 (0.79)	103.62 (12.89)
4 (middle down)			95.08 (0.18)	
1 (left up)		el. dens. 0.01 Å^{-3}	159.83 (0.95)	146.91 (19.25)
4 (middle down)			133.99 (0.21)	
4 (right down)		H-L orbs.[b] 0.02 Å^{-3}	147.36 (0.13)	147.36 (1.06)
4 (right up)	Top and bottom	van der Waals	167.01 (0.24)	170.50 (3.99)
5 (left)			170.12 (0.13)	
5 (right)			174.37 (0.13)	
1 (right up)		Connolly	250.91 (1.19)	228.70 (30.45)
1 (left down)			138.99 (1.16)	
4 (middle up)			196.19 (0.26)	
1 (right up)		el. dens. 0.01 Å^{-3}	316.61 (1.33)	267.96 (49.05)
1 (left down)			318.75 (1.34)	
4 (middle up)			255.46 (0.30)	
5 (left)			224.60 (0.15)	
5 (right)			224.36 (0.15)	
4 (right up)		H-L orbs.[b] 0.02 Å^{-3}	265.54 (0.17)	265.54 (0.55)
1 (left and right)[c]		van der Waals[b]	228.96 (0.66)	228.96 (12.99)
5 (left)	Top	el. dens. 0.06 Å^{-3}	172.11 (0.13)	174.83 (4.03)
5 (right)	Bottom		177.55 (0.13)	
4 (right middle)	Side	van der Waals[b]	105.69 (0.11)	105.69 (0.69)
1 (right down)		Connolly	132.33 (0.86)	122.90 (14.21)
4 (middle)			113.48 (0.20)	
1 (right down)		el. dens. 0.01 Å^{-3}	178.78 (1.00)	177.91 (13.46)
4 (middle)			165.20 (0.24)	
6 (middle up)			189.74 (0.56)	
4 (right middle)		H-L orbs.[b] 0.02 Å^{-3}	200.42 (0.15)	200.42 (4.03)

[a] Not presented in the color plates.
[b] Measured three times to achieve more reliable experimental errors. The surface area presented as $S_i = \langle S\rangle$.
[c] Two-dimensional projection of van der Waals volume of 32 overlapped inhibitors from the training set, drawn applying van der Waals radii for light atoms (excluding S and halogens) on color Plate 1 [26] from a work of Pérez and co-workers [27].

Table 2
Measured (with experimental errors) and calculated geometrical parameters of inhibitor **34** (color Plate 5)

ρ (Å^{-3})	$\langle S\rangle$ (Å^2)	V (Å^3)	H (Å)	L (Å)	D (Å)
0.001	245.96 (3.24)	1350.6	11.71(8)	24.48(8)	2.22
0.01[a]	201.11 (3.48)	854.2	10.20 (8)	23.25(8)	1.59
0.05	172.78 (3.32)	591.4	9.30 (8)	22.36(8)	1.14
0.1	161.53 (3.15)	497.2	8.61 (8)	21.91(8)	0.90
0.2	150.38 (2.66)	413.7	8.20 (8)	21.21(8)	0.57
0.5	136.13 (3.43)	318.3	7.59 (8)	20.89(8)	0.39
0.06[a]	169.90	565.6	–	–	–
0.06536[b]	168.46	553.7[b]	8.96[b]	22.08[b]	1.00[b]
0.001 a.u.[c]	208.62	1450.8	–	–	–
0.002 a.u.[c]	195.90	1277.0	–	–	–

[a] Values for V used for calculation of coefficients of the $V^{1/3}$–log ρ linear equation.
[b] Data used for normalization of D to be $D = 1.00$ Å at van der Waals surface.
[c] In atomic units. Physical boundaries of atoms according to Bader [35].

Table 3
Inhibitor **34**-environment interaction parameters from the F curves

Groups of **34**	F_m (%)	D_m (Å)	I (%)	M_F (%)	D_F (Å)	B (Å)
All groups	100	1.6	100	82	3.1	3.1
H-bond groups	66	2.3	47	45	2.8	1.6
Hydrophobic groups	82	2.7	80	82	3.2	1.7
Aromatic groups	60	3.1	62	56	3.4	1.8
Benzene carbons	29	3.5	28	17	4.0	2.5
Polar groups	82	2.4	70	54	3.0	1.6

Table 4
QSAR data for derivatives of **34**

Molecules	Y_{emp}	Equation for Y_{emp}	X_9	X_{10}	N_{HB}	$N_{5.5}$	$E_{5.5}$ (kJmol^{-1})	E_{10} (kJmol^{-1})	z (kJmol^{-1})
34	9.161	y_{34}	4.0	10	15	440	−78.73	−110.10	−0.18
34[a]	9.161	y_{34}	4.0	10	16	434	−73.67	−105.08	−0.17
54	0	0	0.0	0	8	171	−21.33	−30.08	−0.13
55a[b]	5.091	$y_{34} - (y_1 - y_{10})$	3.0	10	15	402	−69.56	−96.38	−0.17
55b[b]	5.091	$y_{34} - (y_1 - y_{10})$	3.0	10	15	401	−71.44	−84.43	−0.18
56	1.021	$y_{34} - 2(y_1 - y_{10})$	3.0	2	15	363	−62.27	−98.16	−0.017
57	11.227	$y_{34} + 2(y_{31} - y_1) + 2(y_{11} - y_1)$	4.0	18	22	446	−138.07	−169.44	−0.31
58a[b]	7.580	$y_{34} - (y_1 - y_{19})$	3.5	10	15	413	−71.00	−99.46	−0.17
58b[b]	7.580	$y_{34} - (y_1 - y_{19})$	3.5	10	15	413	−70.88	−99.43	−0.17
59	5.999	$y_{34} - 2(y_1 – y_{19})$	3.0	10	15	386	−63.07	−88.69	−0.23
60[c]	11.536	$y_{57} - (y_{11} - y_1)/2 + 2(y_3 - y_1) + (y_{17} - y_1)$	5.0	17	–	525[e]	−153.48[f]	−193.04[f]	−0.29
	11.540[d]	$2.675X_9 - 1.833$							

[a] The case with coordinates of **34** in orientation 2.
[b] In the case when P_1 and P'_1 (**55**), or P_2 and P'_2 (**58**) are not equivalent, both possibilities **a** and **b** were studied.
[c] All values estimated using various schemes.
[d] Regression equation, $r > 0.92$.
[e] The mean from three values predicted using $N_{5.5}$–V, $N_{5.5}$–S and $N_{5.5}$–X_9 regression relationships ($r > 0.97$).
[f] Estimated with respect to **57** as the reference compound: $E_{5.5}(\mathbf{60}) = E_{5.5}$ (**57**) + [$N_{5.5}$(**60**)–$N_{5.5}$(**57**)] $\langle\zeta\rangle$ and $E_{10.0} = 1.165\ E_{5.5} - 14.247$ ($r > 0.98$).

Table 5
Molecular descriptors X_1–X_{14}, S, V and predicted activities for derivatives of inhibitor **34**

Variable[a]	**34**	**54**	**55**	**56**	**57**	**58**	**59**	**60**
X_1	618.777	174.210	528.773	438.526	682.774	568.717	518.656	807.989
X_2	38	18	35	32	54	36	34	56
X_3	38	8	31	24	40	34	30	40
X_4	51	11	43	35	55	46	41	64
X_5	2.705	2.692	2.720	2.742	2.848	2.683	2.658	2.707
X_6 (Å^{-2})	0.05843	0.09804	0.06371	0.07136	0.08548	0.05943	0.06060	0.06947
X_7	30	0	24	18	30	26	22	30
X_8	37	5	30	23	37	33	29	45
X_9	4.0	0.0	3.0	2.0	40	3.5	3.0	5.0
X_{10}	10	4	10	10	18	10	10	17
X_{11}	4.0	0.0	3.0	2.0	40	3.5	3.0	4.0
X_{12} (Å^3)	76.6	59.8	76.6	76.6	118.2	76.6	76.6	114.4
X_{13}	19.723	5.657	18.439	17.059	21.656	17.550	15.133	21.679
X_{14}	60	24	52	44	76	56	52	78
S (Å^2)	650.37	183.59	549.39	448.41	631.75	605.71	561.05	806.09
V (Å^3)	553.7	137.3	464.5	375.3	540.3	509.1	464.5	680.0
Y_{pred}	11.160	−4.883	7.901	4.629	15.636	9.318	7.436	15.740

[a] Molecular descriptors from part I: [10] X_1—relative molecular mass; X_2—π-bonds and lone pairs from heteroatoms; X_3—number of non-hydrogen atoms in planar fragments; X_4—the number of chemical bonds; X_5—the number of valence electrons per atom; X_6–X_2/S; X_7—the number of non-hydrogen atoms in ring systems; X_8—the number of hydrophobic CX_n groups, where $X = H$ or halogen; X_9—the effective number of substituents; X_{10}—the number of potential hydrogen bond donors and acceptors; X_{11}—the effective number of ring substituents; X_{12}—the van der Waals volume of the polar groups; X_{13}—the length of the total aromatic vector including the number of atoms in localized, delocalized and aromatic π-systems, the number of atoms lone pairs, and the number of C atoms in CH_n groups; X_{14}—similar to X_{13}, the total number of non-σ electrons that can be involved in aromatic vectors.

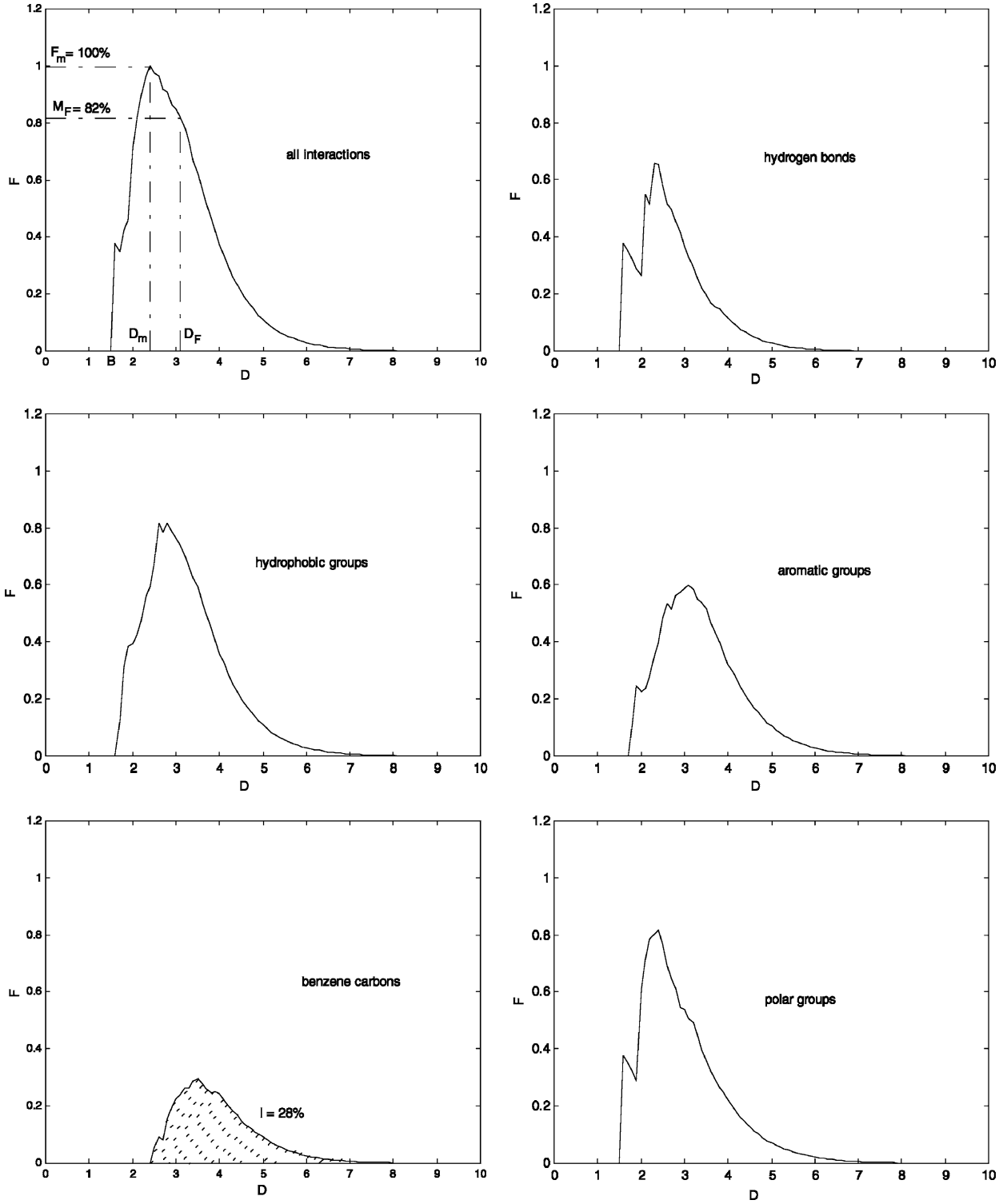


Fig. 2. The function F for inhibitor **34**. Left top: all inhibitor–environment interactions included. Right top: only 31 hydrogen bonds from literature [30] included. Left middle: hydrophobic groups of the inhibitor included (polar group –OH, –NH– and =C=O excluded). Right middle: aromatic groups (benzene rings with their hydrogens and attached carbons) of the inhibitor included. Left bottom: benzene carbons of the inhibitor included. Right bottom: the polar groups of the inhibitor included. D is in Å units. The curve variables B, D_m, F_m, D_F and M_F are defined in the left top picture, and the area under the curve I is presented in the left bottom picture.

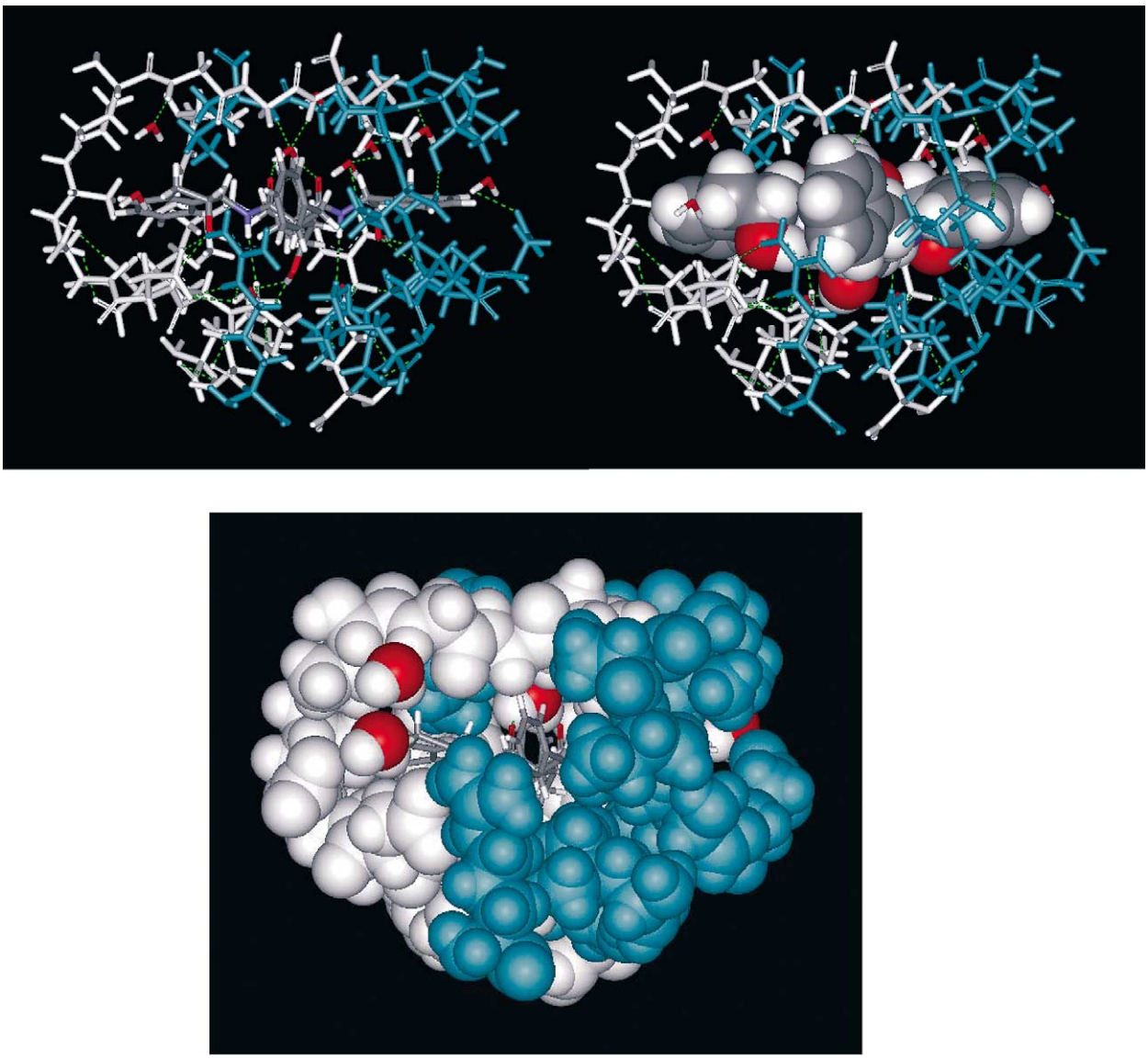
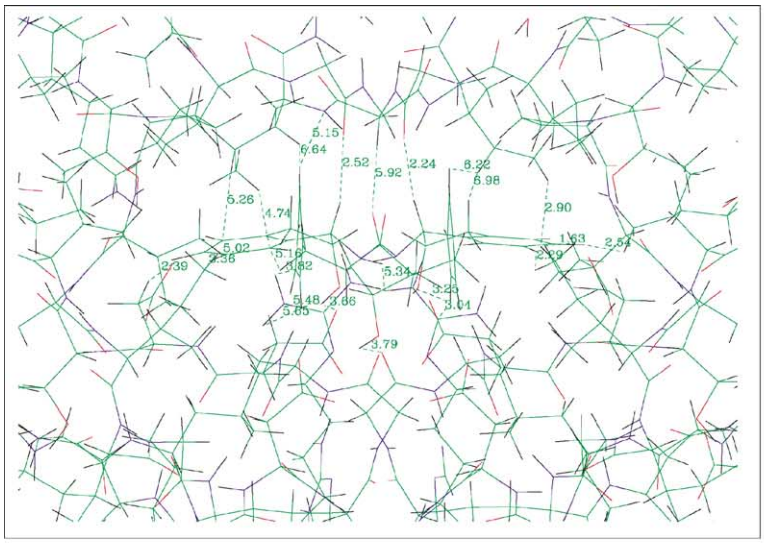
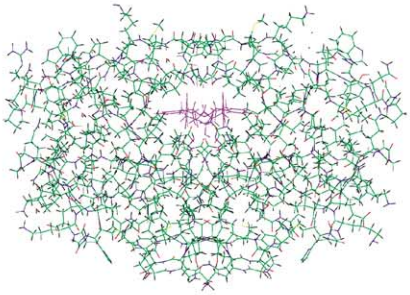


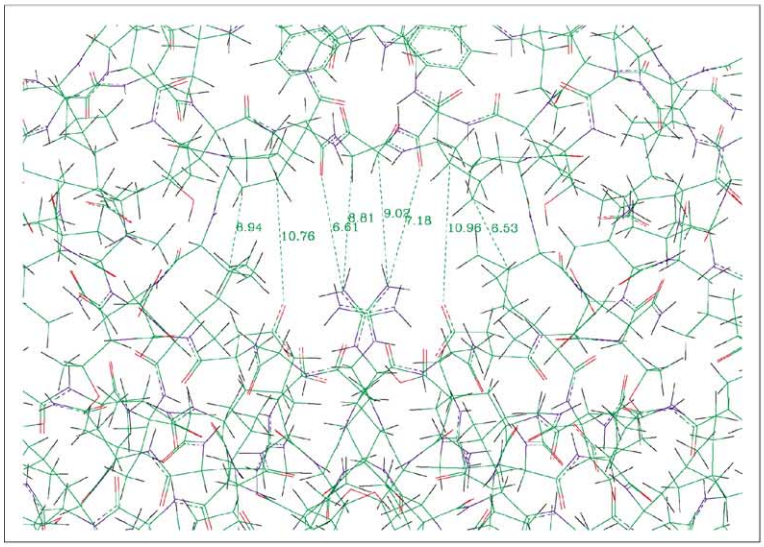
Plate 2. The 29 active site amino-acids (chain A: white, chain B: blue) and 10 water molecules around the inhibitor **34** (yellow) at the cut-off distance 5.5 Å (0.1 Å tolerance). The hydrogen bond network is shown (green).

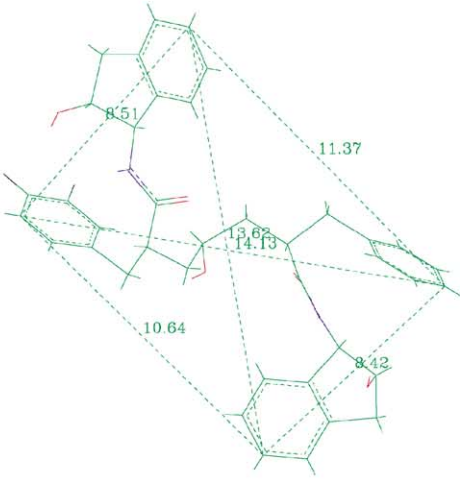
nature of the functional groups. Analysis of color Plate 4 (measuring the Δ thickness of electron density around Connolly volume, see Section 2) exhibits two types of hydrogens. Hydrogens in polar groups are characterized by more diffuse electron density ($\Delta = 1.16(5)$ Å) than hydrogens in hydrophobic groups ($\Delta = 1.02(5)$ Å). The aromatic ring has a characteristic disc shape, but its electron density isosurface, although having the same shape, varies in thickness Δ depending on direction with respect the ring plane. The Δ is minimal in the direction perpendicular to the aromatic ring plane [$\Delta = 0.66(5)$ Å], increases in the direction inclined and is about 45° with respect to the ring plane [$\Delta = 0.87(5)$ Å], and is maximal in the plane of the ring and in between two C–H bonds [$\Delta = 0.97(5)$ Å]. Two trends are observed: (1) Δ, describing aromatic electron density envelope, strongly depends on the distance q parabolically (see Section 2 and Scheme 1 for the definition of q); (2) Δ around C, H and O atoms [0.66 (5), 1.02 (5), 0.76 (5) Å] decreases with the van der Waals radii (1.70, 1.20, 1.52 Å) [23]. This leads to the conclusion that hydrogens result in more diffuse electron density. The hydrogen-deficient/compact groups in organic molecules are aromatic or conjugated rings, which, due to planarity, have relatively high electron density concentrated closely on both sides of the rings' planes. This behavior is not observed for the other functional groups in this work.

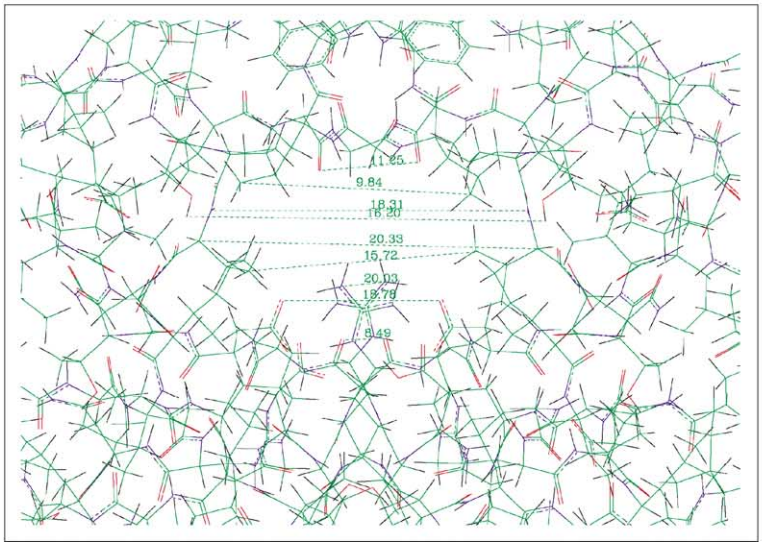
The top and bottom view of **34** is characterized by a linear relationship between volume and its two-dimensional projection. This relationship was established based on the van der Waals (553.7 $Å^3$) and Connolly (725.1 $Å^3$) volumes. The predicted volumes for **34** confirmed the linearity in range 170–230 $Å^2$: the volume enclosed by the 0.06 $Å^{-3}$











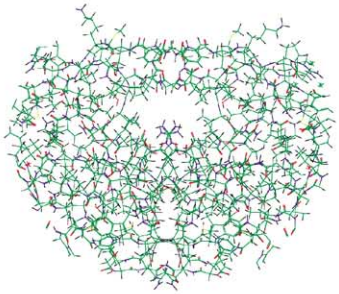


Plate 3. The geometry of the HIV-1 protease active site in free and bound state and of the inhibitor **34** in free state. The atom types are colored in a standard way. The HIV-1 protease in free state (left: middle and bottom; right: bottom): the active site at the binding pockets S_1, S_1', S_2, S_2' has dimensions $\approx$11 Å $\times$ 20 Å $\times$ 20 Å, what is in accordance to those of the free inhibitor **34** (right: middle). The conformational change of the protease due to complexion is obvious (right: top and bottom).

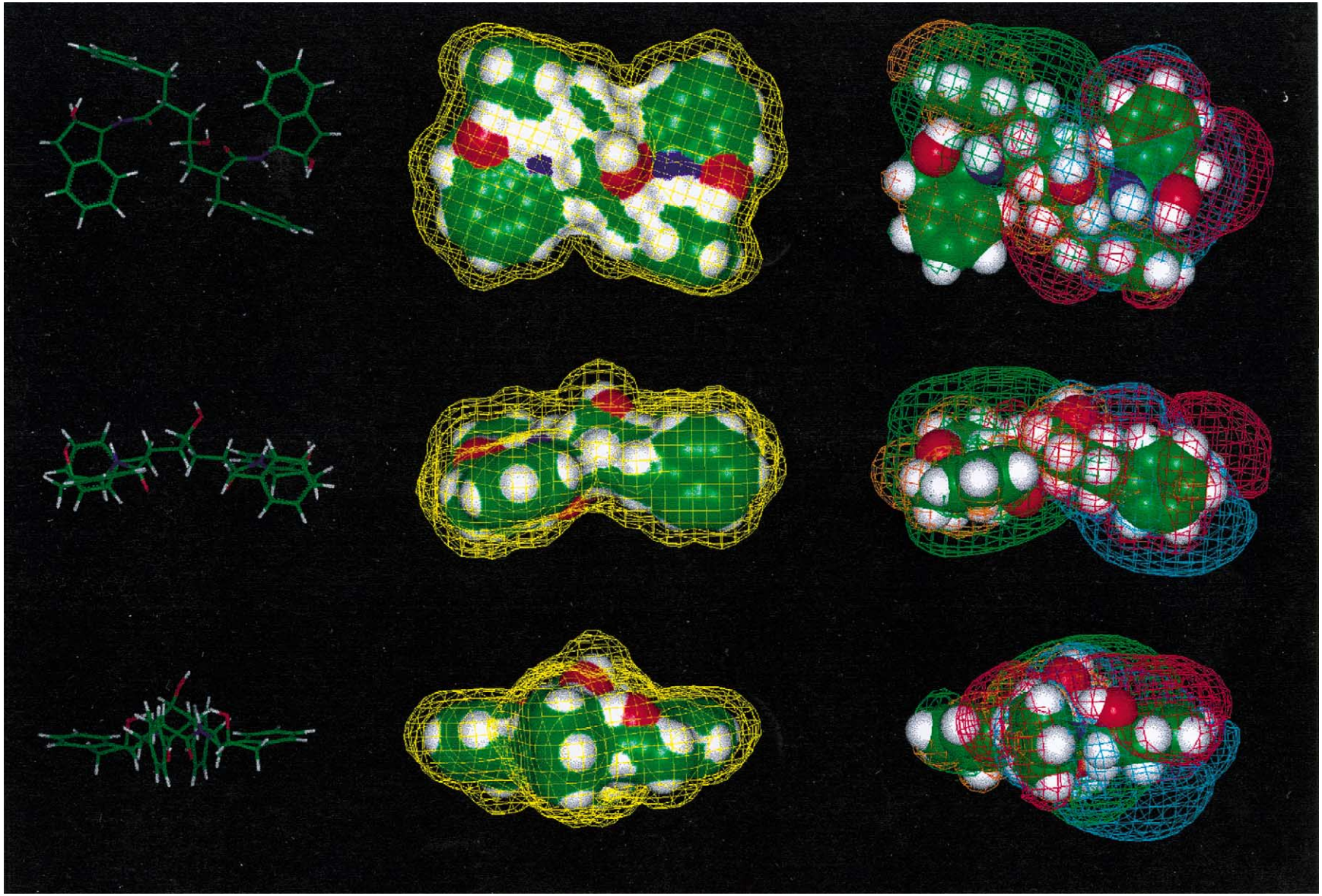


Plate 4. Various views of free inhibitor **34** represented by stick model (left), Connolly surface enclosed into 0.01 Å^{-3} electron density isosurface (middle), and van der Waals model with ±0.02 Å^{-3} HOMO ("+" cyan; "−" red) and LUMO ("+" green; "−" orange) isosurfaces. Atom types are colored in a standard way. The electron density around the central isostere OH group (between the aromatic substituents and oriented upwards, see the stick models) tends to be more diffused than around the other groups in average due to the nature of the hydrogen bond. HOMO and LUMO isosurfaces are even more diffuse than the electron isodensity surface, and do not exhibit C_2 symmetry. Both electron density and HOMO/LUMO isosurfaces point out the significance of mutual penetration of the inhibitor and the enzyme into their "soft" parts.

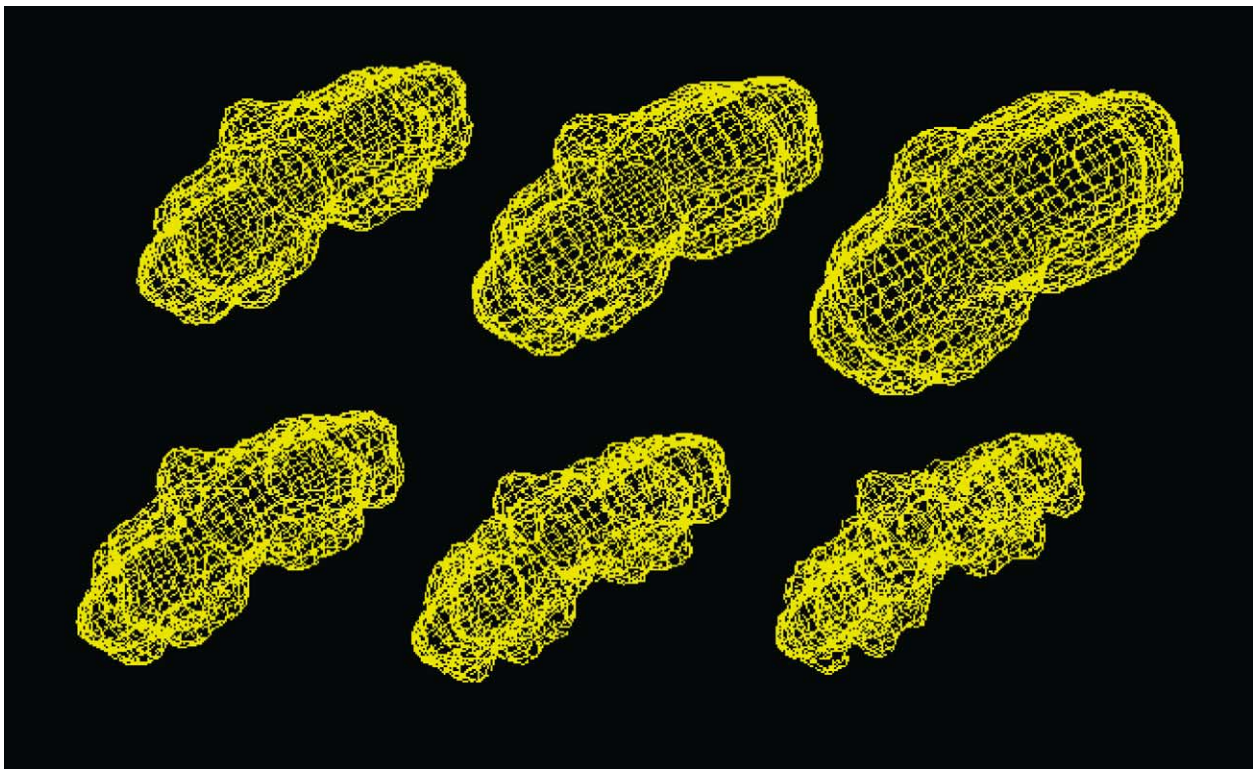


Plate 5. The PM3 electron density isosurfaces (yellow chicken cage) of inhibitor **34**. Top row, from the right to the left: 0.001, 0.01 and 0.05 e Å^{-3}. Bottom row: from the left to the right: 0.1, 0.2, and 0.5 e Å^{-3}. The value 0.001 e Å^{-3}is below the Bader's proposed values 0.007–0.013 e Å^{-3}, while 0.05 e Å^{-3} is close to van der Waals molecular surface.

electron isodensity surface is 566.4 Å^3, close to van der Waals volume. The van der Waals volume of 32 superimposed inhibitors (Table 1) is predicted 725.9 Å^3 from two-dimensional projection of the superimposed set, and the equivalent volume employing the volume increments [25] is 733.7 Å^3. The volumes for 0.01 Å^{-3} electron density isosurface and ±0.02 Å^{-3} HOMO–LUMO isosurface are predicted 840.7 and 834.8 Å^3, respectively. Practically of the same size, these volumes are 1.5 times greater than van der Waals volume, and 1.2 time larger than Connolly volume. These findings point out that molecules must be considered as objects with dimensions produced by van der Waals and even Connolly radii. The linear relationship V–$S^{3/2}$ used instead of V–S yielded 565.6 Å^3 for 0.06 and 854.2 Å^3

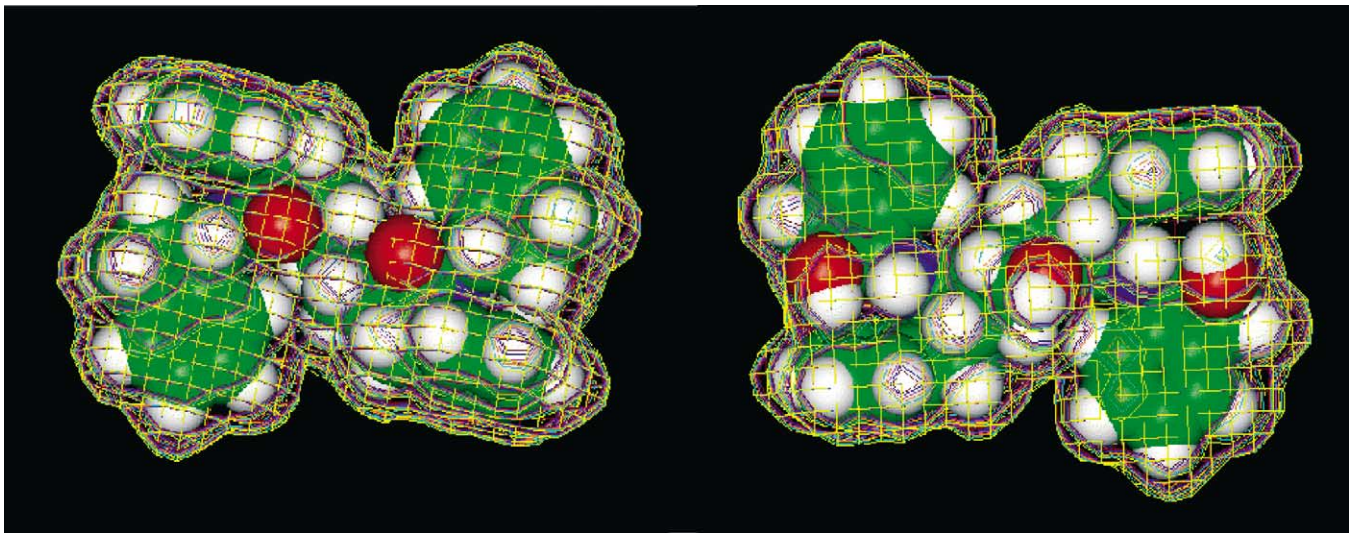


Plate 6. The electron density isosurface (chicken cage) of inhibitor **34** in two views. Various values are represented with colors: yellow—0.06, light blue—0.05, light brown—0.04, pink—0.03, dark blue—0.02, and dark red—0.01 electrons per Å^3.

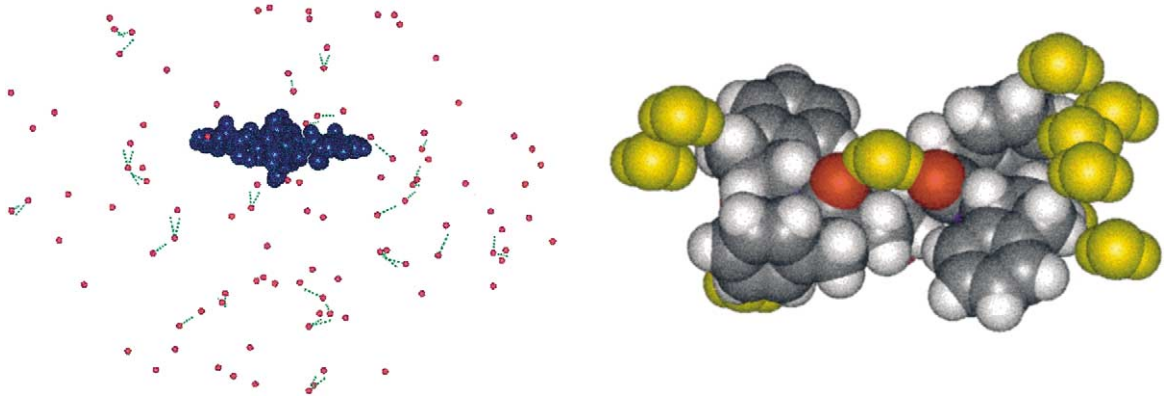


Plate 7. Inhibitor **34** surrounded by water molecules from the crystal structure of HIV-1 protease inhibitor **34** complex (PBD code: 4PHV). Left: all water molecules (red oxygens) are concentrated on the surface and in the holes of the protease. Right: the inhibitor molecule in contact with ten water molecules at cut-off distance 5.5 Å (0.1 Å tolerance). Hydrogen bonds are presented as green dashed lines.

for 0.01 $Å^{-3}$ isosurfaces, 847.3 $Å^3$ for HOMO–LUMO and 725.9 $Å^3$ for the overlapped inhibitors.

3.1.3. Empirical study of inhibitor–environment intermolecular interactions

Table 2 and Fig. 2 summarize the electron density analysis presented in color Plate 5. The new $\langle S \rangle$ value for electron density isosurface 0.01 $Å^{-3}$ is in Table 1. Relative errors of 1–3% are in the same order of magnitude as in Table 1. Predicted volumes for electron isosurfaces of 0.001 and 0.002 a.u. (atomic units), are 2.3–2.6 times greater than van der Waals volume (defined by H and L molecular dimensions) and should be real molecular boundaries according to Bader [35] $\log \rho$ is highly correlated with H and L, showing that electron density has slightly different behavior depending on the distance in a particular direction. This is clearly illustrated by color Plate 6; the ratio L_0/H_0 (2.46) is close to the ratios of variances of the first two PCs when PCA is applied to atomic coordinates of bound (2.83) and free (2.89) **34**. H and L are selected to be molecular dimensions in directions close to the best views (the PCs). All the values in Table 2 were used to obtain F–D curves in Fig. 2 (see Section 2). After 7 Å the curves practically vanish. The choice of 5.5 Å as the molecular boundary was shown to be reasonable. When all environment–inhibitor interactions are considered (see axes in Fig. 2, left up), F is 0.8% of its maximum, and the area I under the curve is 98% of the total area. I is 96.2 and 92.5% for $D = 5.0$ and 4.5 Å, respectively. After 5.5 Å, long-range interactions are practically out of physical possibility for a molecule, so 10.0 Å was used for the cut-off distance. It is worth noting that the counted number of environment atoms inside the cut-off distance D becomes constant at 24.8 Å, which is practically the HIV-1 protease radius (hydrodynamic: 26.7, crystallographic: 26.0 Å) [37]. The inhibitor is placed deeply inside the cavity in such a manner that it reaches the center of the enzyme (color Plates 1 and 3).

Fig. 2 compares inhibitor–environment interactions including specific functional groups of **34**: all groups, those involved in hydrogen bonds, hydrophobic and aromatic species, the benzene carbons as well as the polar groups. Six parameters describe effectively the types of interactions (see Fig. 2, left top and bottom for the definition of the parameters). These include the curve global maximum F_m and its position D_m, the area under the curve I, the weighted mean value M_F of F and its position D_F, and the D for the shortest interaction B. One can notice from Table 3 that there is a significant contribution from hydrogens to the total interactions. The shortest interactions ($B = 1.6$ Å) are the strongest hydrogen bonds. D_m for aromatic groups is 0.4 Å longer than for hydrophobic groups, meaning that aromatic rings participate in interactions with less hydrogen and more non-hydrogen atoms than interactions with all hydrophobic groups. The F curve for aromatic groups reveals their contribution to the total intermolecular interaction. The F curve for benzene atoms is significantly lower than that for all aromatic groups, meaning that most of the interactions including benzene rings are π(benzene)···H(environment) or H(benzene)···X(environment), X being any atom. F curves for H-bonds and the polar groups are similar due to the fact that polar groups are involved mainly in H-bonds (color Plate 7). When these F data (Table 3) are treated with HCA, the resulting dendogram (not shown) has two clusters: H-bonds/polar groups, and aromatic groups/hydrophobic groups. These six parameters can be used as intermolecular interaction descriptors for aromatic systems to characterize aromaticity in intermolecular interactions.

3.2. A priori approach and molecular modeling

The a priori model can be used as a prototype for a more quantitative model by optimizing unrefined molecular geometry, calculating new descriptors or refining a priori variables. Descriptors for a peptidic inhibitor from a frozen protease-frozen inhibitor complex are, in fact, a

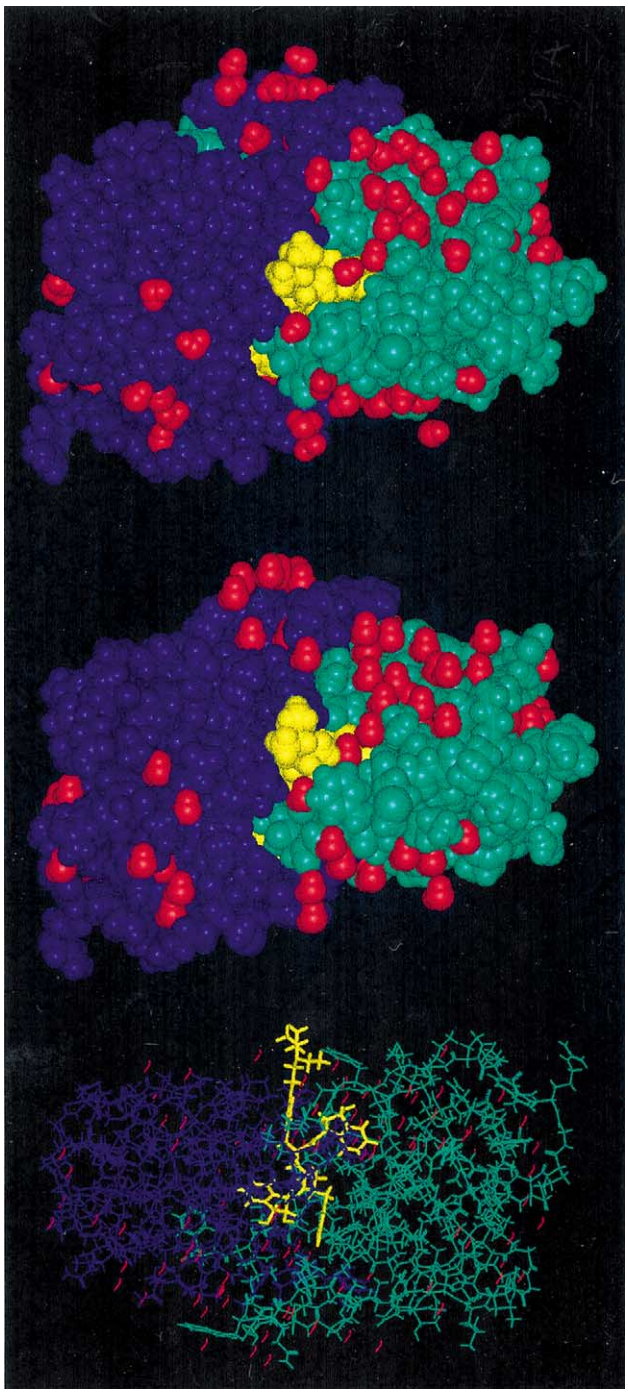


Plate 8. The structure of HIV-1 protease inhibitor **60** complex. Bottom: stick model of the unoptimized complex in top view. The protease chains are shown with different colors, the inhibitor is yellow, and the waters are red. The hook –O–$(CH_2)_2$–N(CH_2CH_3) –CH(CH_3) –CH_2CH_3 in *para*-position of the phenyl of P_1 substituent is shown to be placed out of the active site hole. Middle: van der Waals model of the same complex rotated by 90°. Top: the optimized complex, in the same orientation as in the middle.

priori descriptors. The logical connection between the a priori approach and molecular modeling can be illustrated via QSAR with descriptors from Tables 4 and 5, and the biological activity Y, which can be treated approximately as an additive variable. The variation in structure of P_2'–P_2 substituents can cause systematic changes in Y because all peptidic inhibitors are in extended conformation in the complexes with HIV-1 protease [38]. So the substituents fit into the protease pockets S_2'–S_2 and behave as mutually independent. Consequently, **57** is predicted to be more active than any in the set **1–49**. Pérez et al. [39] and Holloway et al. [32] calculated $E_{10.0} = -88.28$ and $-123.4\,\text{kcal mol}^{-1}$, respectively. $E_{10.0}$ in this work, averaged over the two orientations of **34** is in accordance with these literature results at $-107.59\,\text{kcal mol}^{-1}$. $E_{10.0}$ for **41** is $-91.73\,\text{kcal mol}^{-1}$ by Pérez et al. [38] and $-148.9\,\text{kcal mol}^{-1}$ by Holloway et al. [32]. Since the biological activity is highly correlated with the protease inhibitor interaction energy [27,32,38], $E_{10.0}$ for the most active **57** is meaningful (Table 4). There are small differences in descriptors for two orientations of **34**, for **55a** and **55b**, or **58a** and **58b**, that is a consequence of C_2 symmetry of **34**, of the protease and its active site [27].

A priori descriptors [10] X_{10} and N_{HB} are highly correlated with each other ($r = 0.865$) and slowly increase Y. There are also well-established linear relationships between Y, the effective number of substituents X_9, the number of environment atoms N_{env}, the energies $E_{5.5}$ and $E_{10.0}$, and the van der Waals volume V and surface area S of the inhibitors ($|r| > 0.7$). This explains why the a priori approach performs effectively. Only ζ is uniform in most of the cases, except for **54** (too small), **57** (too many H-bonds), and **59** (distorted polar/apolar ratio in size, number and energy contribution of functional groups). Bigger substituents are involved in many more intermolecular interactions, so the energy of inhibitor-enzyme interaction at any cut-off distance is high due to the uniformity in ζ (Table 4). The size of substituents (as X_9 or another extensive variable) is directly related to biological activity. Pattabiram [40] found that important receptor-drug interactions occur inside the 5.5 Å cut-off. The number of contacts in five HIV-1 protease inhibitor crystal structures (PDB: 1HIV, 1HPV, 1HPX, 1PVI, 1HVJ) [39] at the cut-off 6.5 Å, are linearly dependent on X_9 and X_{11} and their analogues (simple number of ring or all substituents; $r > 0.73$). Zhang et al. [41] showed that 5.5 Å cut-off is the molecular boundary.

How much is the a priori approach accurate in terms of $E_{10.0}$? To estimate the error of $E_{10.0}$, the set of representatitve inhibitors in Table 4 was used. Excluding **54** which does not act as an inhibitor, the interaction energy per environment atom $\langle\zeta\rangle = 0.20\,\text{kcal mol}^{-1}$ and its standard deviation $\text{S.D.}(\zeta) = 0.05\,\text{kcal mol}^{-1}$ gave a relative error of $\text{S.D.}(\zeta)/\langle\zeta\rangle = 24\%$ as a variation of ζ. $\langle N_{env}\rangle = 411$ represents an overall number of environment atoms around the inhibitor at 5.5 Å cut-off. Prediction of N_{env} from linear relationship with descriptor X_9 yields an average deviation of predicted data from measured (Table 3) $\langle\delta\rangle = 5$. So the error in $E_{10.0}$ due to variation of ζ is $\varepsilon_1 = \langle N_{env}\rangle\,\text{S.D.}(\zeta)$, and the error in estimation of N_{env} from X_9 is $\varepsilon_2 = \langle\delta\rangle\text{S.D.}(\zeta)$ yielding the total error $\varepsilon_T = (\varepsilon_1^2 + \varepsilon_2^2)^{1/2} = 19.3\,\text{kcal mol}^{-1}$ or $18\%\langle E_{10.0}\rangle$. The difference between $E_{10.0}$ by Pérez et al. [39]

and Holloway et al. [32] is almost double, 35.1 kcal mol^{-1}, or 40 and 28% of $\langle E_{10.0} \rangle$, respectively. So the a priori approach is comparable to various QSAR software packages.

Color Plate 8 represents the raw (a priori) and MMFF94 optimized structures of the HIV-1 protease inhibitor **60** complex, where **60** is modeled from **57** by adding a lipophilic chain to *para*-positioned OH of the P_1 phenyl, similarly as an aromatic and chain systems are joined in antibiotic actinomycin D [42]. Its peptide hook participates in many weak interactions with DNA. In Table 4 are the calculated molecular descriptors for **60** from the most appropriate equations. There is no significant difference between the raw and optimized complexes (color Plate 8), which justifies the use of the a priori approach. **60** is predicted to be 19 times more active than **41** (the most active in **1–48**). On the other side, biological activities for **34**, **54–60**, predicted using the a priori model I (Table 5) correlate highly with empirically predicted activities (Table 4; $r > 0.93$). For Hansch and co-workers [42] high lipophilicity is a crucial molecular property for good HIV-1 protease inhibitors (C log *P* for **39**, **50**, **51**, **53** ranging from 3.68–5.84). The rescaled log P (with respect to C log *P* for **39** by Hansch et al. [43]) are 3.33 for **57** (many –OH groups), 3.53 when *ortho*-OH on P_1 is replaced by $-(CH_2)_2-NH-CH_2CH_3$, and 4.92 for **60** (in range 3.68–5.84). C_{60}-fullerenes [44] and cyclic ureas [43,45] as HIV-1 inhibitors show how much the lipophilicity is important for protease inhibitors.

3.3. The protease structure–function relationships

HIV-1 protease [38] is a dimer (color Plates 1–3) with an elliptical open-ended cylinder active site (color Plates 1 and 2) formed as well-defined subsites (S_1, S_1', S_2, S_2', S_3, S_3') and not as clear (S_4, S_4', S_5, S_5') subsites containing mostly hydrophobic residues. Interactions with S_2–S_2' are critical for potent binding [46]. The active site can accommodate six–eight amino-acids [37] with its 18 potential H-bond donors/acceptors. Only six–eight donors/acceptors participate in binding the smaller inhibitors [46]. Among **1–49** (Table 3, part I), there are 10 samples (**18**, **20**, **29**, **30**, **38**, **42**, **43**, **45**, **47**, **49**) with this number of hydrogen bonds, and **34** makes over 10 bonds (color Plate 2). The variation in the active site size due to big conformational changes [38] in complexation (color Plate 3) is 853–1566 Å^3 calculated by Liang et al. [47]. All crystal structures of HIV-1 protease–peptidic inhibitor complexes revealed that the bound inhibitors are in an extended conformation where the inhibitor's main chain is extended [38]. This enables the calculation of X_{13}–X_{14} from two-dimensional formulas. The protease's natural substrates are viral *gag* and *gag-pol* polyproteins with mostly hydrophobic groups [48], similarly as **1–48** (Figs. 2 and 3, Part I) and the S sites. Inhibitors saquinavir, ritonavir, indinavir and nelfinavir [43] are good indicators that hydrophobicity is important for anti-HIV activity.

3.4. Enzyme–inhibitor intermolecular interactions

It has been observed [49] that van der Waals radii [50] change from bond to bond, and from conformation to conformation. Bader [49] and Pácios [51] recommended use of contact atomic radii at an electron isodensity of 0.001–0.002 a.u. Besides that, the type of intermolecular interactions is also important. The crystal structure of the enzyme-inhibitor **34** complex [13] is a good example of protease inhibitor molecular complementarity (see color plates). In color Plate 1 polar-polar and hydrophobic–hydrophobic interactions are visible. Polar–hydrophobic interactions are unfavorable [39] and are largely not present. In organic cyrstals, the longest H···H and C···C intermolecular contacts define the molecular boundaries. According to Bader [35] and Pácios [51] these contacts are at 3.1 and 3.8 Å for electron density 0.001 a.u. Mulichack et al. [52] use 4.5 Å, Pattabiram [40] and Zhang et al. [41] use 6.4 Å for the longest C···C contacts, and Stahl and Böhm [53] 7.0 Å. The function *F* yields 5.5 Å. Gavezzotti's atom–atom potential energy minima [31] are at 3.36 and 3.89 Å for H···H and C···C, respectively, and 1.60 Å for the shortest H-bonds. Cut-off distance for long-range interactions is 10.0 Å by Pérez et al. [39] and in this work also ($E_{10.0}$, is ≈95% of the total energy). Weak interactions C–H···O, H···H, C–H···π also stabilize the complex (color Plate 3). Water participates also in weak C–H···O interactions (color Plate 7). It is know that HIV-1 inhibitors should have $M_r > 500$ [55], high lipophilicity ($\log P = 2$–4) [54] and to be similar to *gag-pol* and *gag* proteins (75% aromatic and 50% rings residues at P_1, P_1'). About 150 crystal structures [56,57] of free and complexed HIV-1 protease have confirmed these conditions. Among these inhibitors, five were clinically approved in combination with HIV-1 reverse transcriptase inhibitors [38,49].

4. Conclusions

Molecular graphics analysis on **34** and its derivatives reveals that the interval 1.6–5.5 Å cut-off contains majority of protease-water-inhibitor interactions defining the physical boundary of **34**. Computer-assisted techniques concerning molecular graphics and modeling, illustrated well the chemical background of the a priori approach.

Acknowledgements

The authors acknowledge FAPESP for financial support.

References

[1] N.C. Cohen, The molecular modeling perspective in drug design, in: N.C. Cohen (Ed.), Guidebook on Molecular Modeling in Drug Design, Academic Press, San Diego, 1996, pp. 1–17.

[2] R.E. Hubbard, Molecular graphics and modeling: tools of the trade, in: N.C. Cohen (Ed.), Guidebook on Molecular Modeling in Drug Design, Academic Press, San Diego, 1996, pp. 19–54.
[3] C.D. Selassie, T.E. Klein, Building bridges: QSAR and molecular graphics, in: H. Kubinyi (Ed.), 3D QSAR in Drug Design: Theory, Methods and Applications, Kluwer Academic Publisher, Dordrecht, The Netherlands, 2000, pp. 257–275.
[4] H.P. Klung, L.E. Alexander, X-ray Diffraction Procedures for Polycrystalline and Amorphous Materials, 2nd ed., Wiley, New York, 1974.
[5] R. Kiralj, M.M.C. Ferreira, QSAR of progestogens: use of a priori and computed molecular descriptors and molecular graphics, J. Chemometr., Quant. Struct.-Act. Relat., in press.
[6] M.M.C. Ferreira, R. Kiralj, Molecular graphics-structural and molecular graphics-descriptors in a QSAR study of 17-α-acetylprogesterones, J. Braz. Chem. Soc., Web available since 19 November 2002.
[7] R. Kiralj, B. Kojic-Prodic, I. Piantanida, M. Zinic, Crystal and molecular structures of diazapyrenes and a study of π···π interactions, Acta Cryst. B55 (1999) 55–69.
[8] M.M.C. Ferreira, R. Kiralj, in: Proceedings of the 11th Simpósio Brasileiro de Química Teórica, Caxambu, MG, Brazil, 18–21 November 2001, p. 227.
[9] K.R. Beebe, R.J. Pell, M.B. Seasholtz, Chemometrics: A Pratical Guide, Wiley, New York, NY, 1998.
[10] M.M.C. Ferreira, R. Kiralj, A priori molecular descriptors in QSAR: a case of HIV-1 protease inhibitors. I. Chemometrics, J. Mol. Graphs. Mod. 21 (2003) 435–448.
[11] Insight II, Version 95.0, 1995, Biosym/MSI Inc., San Diego, CA, USA.
[12] WebLab Viewer, Version 2.01, 1997, Molecular Simulations Inc., San Diego, CA, USA.
[13] R. Bone, J.P. Vacca, P.S. Anderson, M.K. Holloway, X-ray crystal structure of the HIV-1 protease complex with L-700,417, an inhibitor with pseudo C_2 symmetry, J. Am. Chem. Soc. 113 (1991) 9382–9384.
[14] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 267 (1997) 727–748, http://www.ccdc.cam.ac.uk/prods/gold/gold.html [accessed on 19 February 2002].
[15] The Protein Data Bank (PDB), The Research Collaboratory for Structural Bioinformatics, Rutgers, NJ. http://www.rcsb.org/pdb/ [accessed on 18 February 2002].
[16] J.J.P. Stewart, MOPAC, Version 6.0, 1990, Frank Seiler Research Laboratory at Air Force Academy, Colorado Springs, Colorado.
[17] J.J.P. Stewart, Reply to comments on a comparison of am1 with the recently developed PM3 method, J. Comput. Chem. 11 (1990) 534–544.
[18] J.J.P. Stewart, Optimization of parameters for semiempirical methods. I. Method, J. Comput. Chem. 10 (1989) 209–220.
[19] J.J.P. Stewart, Optimization of parameters for semiempirical methods. II. Applicatons, J. Comput. Chem. 10 (1989) 221–264.
[20] M.M.C. Ferreira, R. Kiralj, New methods for finding the best orientation in molecular graphics, unpublished results.
[21] C.K. Clarke, Geography 115A/115AL: Introductory Air Photo Interpretation and Remote Sensing, Lecture 3: Scale and Area Measurement, Department of Geography, University of California, Santa Barbara, CA, USA. http://www.geog.ucsb.edu/~jeff/115a/ [accessed on 6 February 2002].
[22] A.H. Robinson, Elements of Cartography, 2nd ed., Wiley, New York, 1966, p. 48.
[23] PLATON, A Multipurpose Crystallographic Tool, Version 31000, 2000, Spek, A.L. Utrecht University, Utrecht, The Netherlands.
[24] F.K. Allen, O. Kennard, D.G. Watson, L. Brammer, G. Orpen, R. Taylor, Tables of bond lengths determined by X-ray and neutron diffraction. Part 1. Bond lengths in organic compounds, J. Chem. Soc., Perkin Trans. 2 (1987) S1–S19.
[25] A. Gavezzotti, Molecular packing and correlations between molecular and crystal properties, in: H.-B. Bürgi, J.D. Dunitz (Eds.), Structure Correlation, vol. 2, VCH Publishers, Weinheim, 1994, pp. 509–542.
[26] M. Pastor, C. Pérez, F. Gago, Simulation of alternative binding modes in a structure-binding based QSAR study of HIV-1 protease inhibitors, J. Mol. Graphics Mod. 15 (1997) 389.
[27] M. Pastor, C. Pérez, F. Gago, Simulation of alternative binding modes in a structure-binding based QSAR study of HIV-1 protease inhibitors, J. Mol. Graphics Mod. 15 (1997) 363–371.
[28] L.F. Pácios, Analytical density-dependent representation of Hartree-Fock atomic potentials, J. Comp. Chem. 14 (1993) 410–421.
[29] R. Bone, J.P. Vacca, P.S. Anderson, M.K. Holloway, X-ray crystal structure of the HIV-1 protease complex with L-700,417 an inhibitor with pseudo C_2 symmetry, J. Am. Chem. Soc. 113 (1991) 9382–9384.
[30] R. Kiralj, M.M.C. Ferreira (Eds.), F77 Routines for Study of Inhibitor–Enzyme Interactions, Instituto de Química, Universidade Estadual de Campinas, Campinas, SP, Brazil, 2000.
[31] A. Gavezzotti, Are crystal structures predictable? Acc. Chem. Res. 27 (1994) 309–314.
[32] 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, A priori prediction of activiy for HIV-1 protease inhibitors employing energy minimization in the active site, J. Med. Chem. 38 (1995) 305–317.
[33] Titan, Version 1, Wavefunction Inc., Irvine, CA, 2000.
[34] T.A. Halgren, Merck molecular force field. 1. Basis, form, scope, parameterization, and performance of MMFF94, J. Comp. Chem. 17 (1996) 490–519.
[35] W.F. Bader, Atoms in Molecules, Clarendon Press, Oxford, 1990, p. 432.
[36] M.R. Spiegel, Schaum's Outline of Theory and Problems of Theoretical Mechanics with Introduction to Langrange's Equation and Hamilton Theory, McGraw-Hill, New York, 1967, p. 255.
[37] A.J. Kungl, N.V. Visser, A. von Hoek, A.J.W.G. Visser, A. Billich, A. Schilk, H. Gstach, M. Auer, Time-resolved fluorescence anisotropy of HIV-1 protease inhibitor complexes correlated with inhibitor activity, Biochemistry 37 (1998) 2778–2786.
[38] A. Wlodawer, J. Vondrasek, Inhibitors of HIV-1 protease: a major success of structure-assisted drug design, Annu. Rev. Biophys. Biomol. Struct. 27 (1998) 249–284.
[39] C. Pérez, M. Pastor, A.R. Ortiz, F. Gago, Comparative binding energy analysis of HIV-1 protease inhibitors: incorporation of solvent effects and validation as a powerful tool in receptor-based drug design, J. Med. Chem. 41 (1998) 836–852.
[40] N. Pattabiram, Occluded molecular surface analysis of ligand-macromolecule contacts: application to HIV-1 protease inhibitor complexes, J. Med. Chem. 42 (1999) 3821–3834.
[41] C. Zhang, G. Vasmatzis, J.L. Cornette, C. DeLisi, Determination of atomic desolvation energies from the structures of crystallized proteins, J. Mol. Biol. 267 (1997) 707–716.
[42] C.Y. Lian, H. Robinson, A.H.-J. Wong, J. Am. Chem. Soc. 118 (1996) 8791–8801.
[43] R. Garg, S.P. Gupta, H. Gao, M.S. Baby, A.K. Debnath, C. Hansch, Comparative quantitative structure–activity relationship studies on anti-HIV drugs, Chem. Rev. 99 (1999) 3526–3601.
[44] S.H. Friedman, D.L. DeCamp, R.P. Sijbesma, G. Srdanov, F. Wudl, G.L. Kenyon, Inhibition of the HIV-1 protease by Fullerene derivatives: model building studies and experimental verification, J. Am. Chem. Soc. 115 (1993) 6506–6509.
[45] A.K. Debnath, Three-dimensional quantitative structure–activity relationship study on cyclic urea derivatives as HIV-1 protease

inhibitor: application of comparative molecular field analysis, J. Med. Chem. 42 (1999) 249–259.
[46] E.A. Lunney, S.E. Hagen, J.M. Domagola, C. Humblet, J. Kosinski, B.D. Tait, J.S. Warmus, M. Wilson, D. Ferguson, D. Hupe, P.J. Tummino, E.T. Baldwin, T.N. Bhat, B. Liu, J.W. Erickson, A novel nonpeptide HIV-1 protese–inhibitor: elucidation of the binding mode and its application in the design of related analogs, J. Med. Chem. 37 (1994) 2664–2677.
[47] J. Liang, H. Edelsbrunner, C. Woodward, Anatomy of protein pockets and cavities—measurement of binding site geometry and implications for ligand design, Protein Sci. 7 (1998) 1884–1897.
[48] M. Sakurai, S. Higashida, M. Sugano, H. Handa, T. Komai, R. Yagi, T. Nishigaku, Y. Yabe, Studies of human immunodeficiency virus type 1 (HIV-1) protease inhibitors. III. Structure–activity relationship of HIV-1 protease inhibitors containing cyclohexylalanylalanine hydroxyethane dipeptide isostere, Chem. Pharm. Bull. 42 (1994) 534–540.
[49] L.M. Rellick, W.J. Becktel, Comparison of van der Waals and semiempirical calculations of the molecular volumes of small molecules and proteins, Biopolymer 42 (1997) 191–202.
[50] A. Bondi, J. van der Waals, Volumes and radii, Chem. Phys. 68 (1964) 441–451.
[51] L.F. Pácios, Atomic radii Scales and electron properties deduced from the charge density, J. Comp. Chem. 16 (1995) 133–145.
[52] A.M. Mulichack, J.O. Hui, A.G. Tomasselli, R.L. Heinrikson, K.A. Curry, C.-S. Tomich, S. Thaisrivongs, T.K. Sawyer, K.D. Watenpaugh, The crystallographic structure of the protease from human immunodeficiency virus type 2 with two synthetic peptide transition state analog inhibitors, J. Biol. Chem. 268 (1993) 13103–13109.
[53] M. Stahl, H.-J. Böhm, Development of filter functions for protein–ligand docking, J. Mol. Graph. 16 (1998) 121–132.
[54] S.S. Abdel-Meguid, B.W. Metcalf, T.J. Carr, P. Demarsh, R.L. DesJarlais, S. Fisher, D.W. Green, L. Ivanoff, D.M. Lambert, K.H.M. Murthy, S.R. Petterway Jr., W.J. Pitts, T.A. Tomaszek Jr., E. Winborne, B. Zhao, G.B. Dreyer, T.D. Meek, An orally bioavailable HIV-1 protease inhibitors containing and imidazole-derived peptide bond replacement: crystallographic and pharmacokinetic analysis, Biochemistry 33 (1994) 11671–11677.
[55] G. Klebe, Structure correlation and ligand/receptor binding, in: H.-B. Bürgi, J.D., Dunitz (Eds.), Structure Correlation, vol. 2, VCH, Weinheim, Germany, 1994, pp. 543–603.
[56] HIV Protease Database, National Cancer Institute at Frederick, Frederick, NW. http://www.ncifcrf.gov/HIVdb/ [accessed on 19 February 2002].
[57] Database for anti-HIV compounds, National Institute for Allergy and Infectious Deseases, Bethesda, MA, US. http://www.niaid.nih.gov/daids/dtpdb/ [accessed on 19 February 2002].