

# Binding of novel fullerene inhibitors to HIV-1 protease: insight through molecular dynamics and molecular mechanics Poisson–Boltzmann surface area calculations

Haralambos Tzoupis · Georgios Leonis ·  
Serdar Durdagi · Varnavas Mouchlis ·  
Thomas Mavromoustakos · Manthos G. Papadopoulos

Received: 10 June 2011 / Accepted: 14 September 2011 / Published online: 4 October 2011  
© Springer Science+Business Media B.V. 2011

**Abstract** The objectives of this study include the design of a series of novel fullerene-based inhibitors for HIV-1 protease (HIV-1 PR), by employing two strategies that can also be applied to the design of inhibitors for any other target. Additionally, the interactions which contribute to the observed exceptionally high binding free energies were analyzed. In particular, we investigated: (1) hydrogen bonding (H-bond) interactions between specific fullerene derivatives and the protease, (2) the regions of HIV-1 PR that play a significant role in binding, (3) protease changes upon binding and (4) various contributions to the binding free energy, in order to identify the most significant of them. This study has been performed by employing a docking technique, two 3D-QSAR models, molecular dynamics (MD) simulations and the molecular mechanics Poisson–Boltzmann surface area (MM–PBSA) method.

**Electronic supplementary material** The online version of this article (doi:10.1007/s10822-011-9475-4) contains supplementary material, which is available to authorized users.

H. Tzoupis · G. Leonis (✉) · T. Mavromoustakos ·  
M. G. Papadopoulos (✉)  
Institute of Organic and Pharmaceutical Chemistry, National  
Hellenic Research Foundation, 48 Vas. Constantinou Ave,  
11635 Athens, Greece  
e-mail: gleonis@eie.gr

M. G. Papadopoulos  
e-mail: mpapad@eie.gr

H. Tzoupis · V. Mouchlis · T. Mavromoustakos  
Chemistry Department, National and Kapodistrian University  
of Athens, Panepistimioupolis Zographou, 15771 Athens, Greece

S. Durdagi  
Department of Biological Sciences, Institute of Biocomplexity  
and Informatics, University of Calgary, 2500 University Drive,  
Calgary, AB T2N 1N4, Canada

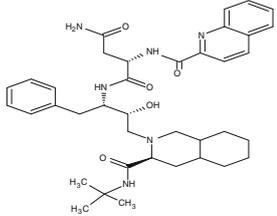
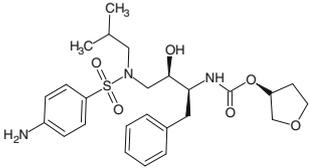
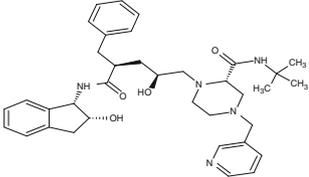
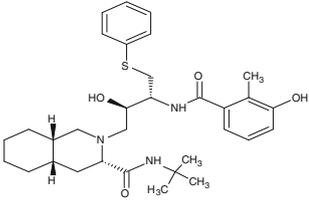
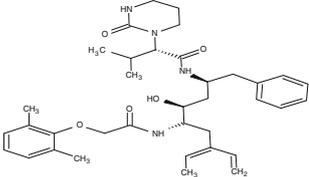
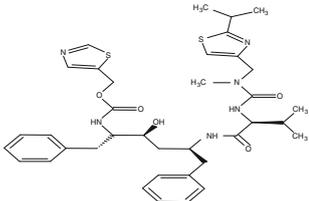
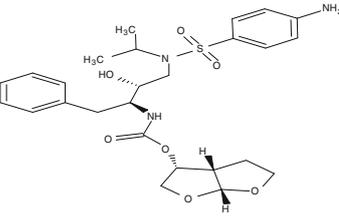
Our computed binding free energies are in satisfactory agreement with the experimental results. The suitability of specific fullerene derivatives as drug candidates was further enhanced, after ADMET (absorption, distribution, metabolism, excretion and toxicity) properties have been estimated to be promising. The outcomes of this study revealed important protein–ligand interaction patterns that may lead towards the development of novel, potent HIV-1 PR inhibitors.

**Keywords** HIV-1 PR · QSAR · Molecular dynamics ·  
MM–PBSA · Fullerenes

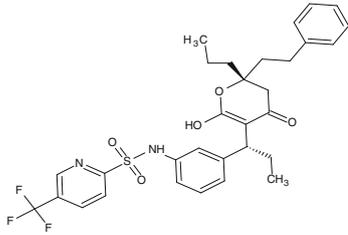
## Introduction

Acquired Immunodeficiency Syndrome (AIDS) has been a major pandemic threat, with more than 33.4 million people infected worldwide, over the last 30 years [1]. Human immunodeficiency virus (HIV) that causes AIDS has been proven to be a very resilient microorganism against various treatment regimens. The mutations appearing in the genome of the virus are creating resistant strains to different drugs, making the discovery of new active compounds more pressing [2]. The major targets for HIV inhibition are two enzymes which control the reverse transcription of the viral RNA (reverse transcriptase, HIV RT), and the partition of the Gag-Pol polypeptide to produce the viral proteins (HIV-1 PR) [3, 4]. The latter belongs to the family of retroviruses and is a symmetric, homodimeric, aspartic protease. It is constituted of two identical polypeptide chains with 99 amino acids each. The active site consists of a catalytic triad in each chain of the dimer with amino acid sequence Asp–Thr–Gly in positions 25, 26, 27 (chain A) and 25', 26', 27' (chain B) [3]. Aspartic acid in positions

**Table 1** Comparison of calculated with experimental binding energies of commercially available HIV-1 PR inhibitors

Name	Molecular structure	Binding energy (kcal mol <sup>-1</sup> )	
		Calculated	Experimental
Saquinavir		-11.37	-14.24 <sup>a</sup>
Amprenavir		-9.83	-13.04 <sup>b</sup>
Indinavir		-11.29	-13.24 <sup>c</sup>
Nelfinavir		-11.18	-12.34 <sup>d</sup>
Lopinavir		-12.69	-16.13 <sup>c</sup>
Ritonavir		-11.51	-14.84 <sup>e</sup>
Darunavir		-12.57	-15.59 <sup>f</sup>

**Table 1** continued

Name	Molecular structure	Binding energy (kcal mol <sup>-1</sup> )	
		Calculated	Experimental
Tipranavir		-11.49	-14.36 <sup>g</sup>
		MRE <sup>h</sup>	0.19

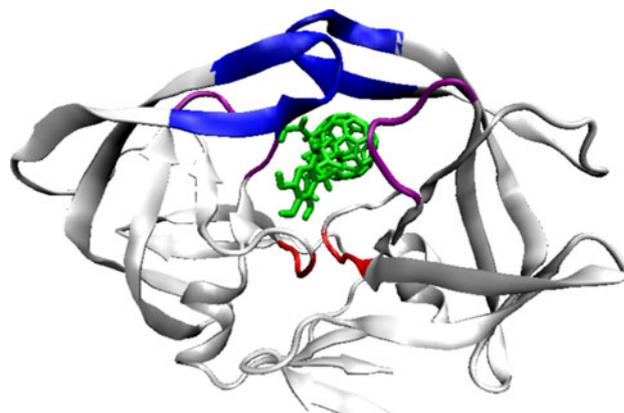
<sup>a</sup> Ref. [48]<sup>b</sup> Ref. [69]<sup>c</sup> Ref. [70]<sup>d</sup> Ref. [71]<sup>e</sup> Ref. [72]<sup>f</sup> Ref. [73]<sup>g</sup> Ref. [74]<sup>h</sup> Relative error (RE) =  $|V_i - V_e|/V_e$ Mean relative error (MRE) =  $1/n \sum(\text{RE})$  $V_i$  = predicted value (CoMFA or CoMSIA) or calculated values (docking);  $V_e$  = experimental value;  $n$  = population size

25/25' catalyzes the hydrolysis of the substrate, which takes place at the cleavage site, by electron transfer [5]. The active site of the protease is covered by two glycine-rich  $\beta$ -hairpins (flaps) that are implicated in modulating the entrance of ligands into the binding cavity. Theoretical and experimental observations suggest that the flaps (residues 44–55 and 44'–55') belong to the most flexible part of the apo protease, appearing in “open”, “semi-open” and “closed” forms, while in the case of bound HIV-1 PR they adopt primarily a closed conformation [6].

Most of the commercially available drugs are peptide analogues, which mimic the natural substrate (Gag-Pol) of the protease (Table 1 and references herein). Inhibition of the protease leads to the production of immature viruses and thus the viral load in patients is decreased [7].

Over the years, fullerene-based analogues have been used as potential enzyme inhibitors [8]. Two pioneering works in this area from Friedman et al. [9] and Toniolo et al. [10] showed the fitting of C<sub>60</sub> into the HIV-1 PR cleft and the possible interactions with the residues of the protein, using molecular simulations and experimental methods. The binding of C<sub>60</sub> with a number of proteins (e.g. HIV-1 PR, cysteine and serine proteases, tubulin, and acetylcholinesterase) has been studied and evidence has been provided that they are a good match [10]. Figure 1 shows a representation of a fullerene analogue (compound 14, investigated in the present study) inside the HIV-1 protease. The fullerene core is hydrophobic, making it

almost impossible to interact with polar solvents such as water [11]. Appropriately designed side chains attached to the fullerene, not only do they increase its solubility in polar solvents, but they also enhance interactions between the compound and biological macromolecules, such as HIV-1 PR [11]. Marchesan et al. [12] have reported cationic fullerenes as potential HIV-1 PR inhibitors. Addition of cationic groups, such as ammonium salt groups,



**Fig. 1** Representation of fullerene compound 14 (green) inside the HIV-1 PR active site cavity. Active site residues (Asp25-Thr26-Gly27, Asp25'-Thr26'-Gly27', red), flap residues (44–55/44'–55', blue), and the solvent-accessible region involving residues 79–83/79'–83' (purple) are also displayed

improves greatly the solubility of the fullerene in polar solvents. Zhang et al. [13] have studied the inhibitory action of fullerene azido analogues derived by appropriately modified fragments of AZT (3'-azido-3'-deoxythymidine, Zidovudine), a known anti-HIV drug. Lee et al. [14] explored the nature and strength of the interactions between fullerene derivatives and the active site residues of the protein. A thorough understanding of the interactions between HIV-1 PR and the inhibitor may facilitate the design of more potent drugs. Troshina et al. [15] have synthesized fullerene derivatives with improved water solubility that appear to be potent HIV-1 PR inhibitors, indicating strong interactions with the active site residues. Furthermore, Promsri et al. [16] have studied the molecular and electronic properties of fullerene-based compounds used as HIV-1 PR inhibitors. Finally, Papadopoulos and co-workers combined various computational approaches to propose a series of novel fullerene inhibitors against HIV-1 protease [17–19].

The first goal of this work was to design a series of novel and promising fullerene-based inhibitors for HIV-1 PR. Thus, we studied the efficiency of two design strategies. The first was based on the results of Comparative Molecular Similarity Indices Analysis (CoMSIA) model, while the second strategy used appropriately modified fragments of the drug saquinavir (Table 1) as side chains to the fullerene core.

The second essential goal of this work was to select specific fullerene derivatives via the above procedure and to analyze their binding modes with HIV-1 PR. Thus, we investigated:

- The H-bond patterns between each ligand and HIV-1 PR.
- The regions of HIV-1 PR, which are of particular importance for the binding.
- The most significant contributions to the binding free energy.

For this study we employed a docking technique, 3D-QSAR models, all-atom unrestrained MD simulations and the MM-PBSA method. Finally, all saquinavir-related fullerenes have been subjected to an analysis regarding their pharmacological activity as expressed by ADMET properties. Many promising drug candidates (in terms of binding affinity estimations, suitable structural features and other predictions) eventually fail in clinical trials mainly due to poor ADMET properties [20]. Prediction of such properties before any attempt to experimentally test a potential drug candidate may be extremely useful, timesaving and cost-efficient. Our results provided significant insight into the binding mode of HIV-1 PR and may help towards the development of improved medications.

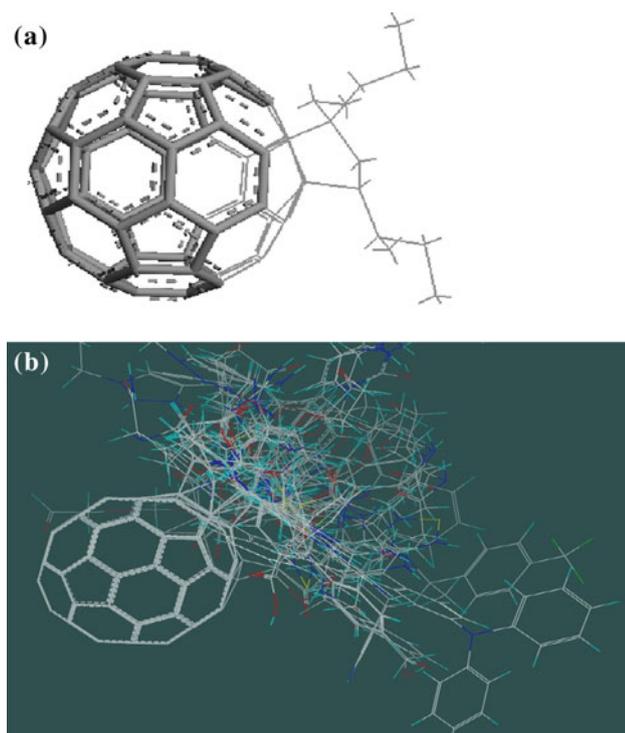
## Methods

### Geometry optimizations and molecular docking simulations

The ArgusLab software [21] was used for the design and the initial optimization of the compounds. Geometry optimization was performed with a unified force field [22]. 1000 steps of steepest descent minimization were followed by 1000 steps with Broyden-Fletcher-Goldfarb-Shanno [23] (BFGS) search, which uses an approximate Hessian matrix. The convergence criterion was set to  $10^{-2}$  kcal mol<sup>-1</sup>. For the docking simulations, the grid resolution was set at 0.400 Å, with a population size of 50. During the simulations, the protease was kept rigid, while the ligand was considered flexible. The active site was directly defined by selecting residues Asp25/25', Thr26/26' and Gly27/27'. The ligands were allowed to move freely in the vicinity of the active site, which was centered in a cubic box of 20 Å × 20 Å × 20 Å. Since genetic algorithms allow the construction of a random statistical initial population, the elitism for the calculations was set to 5. The mutation and crossover rate for the docking simulations were set to 0.2 and 0.8, respectively, while the convergence criterion was set to 1.00 kcal mol<sup>-1</sup>.

### Pharmacophore modeling

For the creation of the 3D-QSAR, we employed CoMFA and CoMSIA. In the development of these models we used 51 derivatives as a training set and calculated their binding energies (Table S1). The compounds within the training set are a representative sample of structures with wide range of binding energies, which vary from high to low values, as required by the criterion for good diversity in the data set [24]. From the 51 compounds of the training set, 46 derivatives have been synthesized for various research subjects (Table S1, footnote a) and 5 compounds have been computationally designed, aiming at the maximization of their binding affinity against HIV-1 PR (Table S1, footnote b). The choice of the side chains was based on the properties of groups such as aromatic rings (e.g. compound **S18**), and hydrogen bond donors or hydrogen bond acceptors, such as -C=O, -NH<sub>2</sub>, or -OH groups (e.g. compound **S22**). The compounds in the training set were aligned as shown in Fig. 2. The calculated binding energies were converted to binding affinities using the formula  $\Delta G = -RT \ln K_i$  [17], after the assumption that the value of  $K_i$  is equal to EC<sub>50</sub> [25]. The cutoff point for both steric and electrostatic contributions was set at 30 kcal mol<sup>-1</sup>. CoMFA QSAR models consider only steric and electrostatic factors, while the CoMSIA models take additionally into account hydrogen bonding and hydrophobic interactions. The models were built using the Tripos CoMFA/



**Fig. 2** **a** The solid cylinder structure of the fullerene core used for alignment in the CoMFA/CoMSIA models, for all compounds in the training set, and **b** a representation of the aligned compounds

CoMSIA module of SYBYL 8.0 molecular modeling package [26]. Finally, a set of available experimental binding energies has been used to confirm the adequacy of our results (test set, Table S2).

#### Molecular dynamics simulations in water

MD simulations in explicit solvent were performed using the SANDER [27] module under the AMBER 9 software package [28]. The high resolution HIV-1 PR crystal structure was obtained from the Protein Brookhaven Databank (PDB code: 1AID) [29]. Crystal water molecules and the ligand (THK inhibitor) were removed before the addition of missing hydrogen atoms, using the tLEaP module of AMBER. Due to the absence of a fullerene-bound crystal structure of HIV-1 PR, the choice of 1AID can be rationalized by the fact that THK is size-comparable to the fullerene. Several studies have considered the protonation state of the aspartic acid residues on the active site of HIV-1 PR [30–34]. Since it is generally accepted that an unprotonated active site is not associated with effective binding, the possibility of either a monoprotonated or a diprotonated active site is prevalent. In our case, the active site was considered monoprotonated (Asp25), according to previous experimental and theoretical studies that support this choice [30, 32, 34]. The force field ff99SB [35] was used to obtain

atomic partial charges and parameters for the protease. Fullerene structures were constructed with the ANTECHAMBER module (using the general GAFF force field, with AM1-BCC atomic partial charges) [36–38]. Explicit solvation was performed using tLEaP after neutralizing the systems. Each complex was solvated using the TIP3P water model [39] ( $\sim 7000$  water molecules added) and truncated octahedral periodic boundary conditions were applied, with a cutoff distance of 8 Å. The long range electrostatic interactions were calculated using the Particle Mesh Ewald (PME) method [40]. The starting step was the minimization of the protein–ligand complex over 2000 cycles with a cutoff of 25 Å. For the first 1000 steps the steepest descent method was used, while for the next 1000 steps the conjugate gradient technique was employed. The next procedure involved the gentle heating of the system under constant volume over 100 ps, with the gradual increase of the temperature from 0 to 300 K. The SHAKE algorithm [41] was applied to constrain all bond lengths involving hydrogen atoms to their equilibrium distance, thus allowing to use a 2 fs time step. The Langevin thermostat with a collision frequency of  $2 \text{ ps}^{-1}$  has been used to keep the temperature constant [42]. An 100 ps constant-pressure equilibration followed, to observe the gradual increase of the density, which converged after  $\sim 40$  ps. The MD simulation was next initiated using a Langevin dynamics temperature scaling with a collision frequency of  $2 \text{ ps}^{-1}$ , for 20 ns (time step: 2 fs). During the MD simulations, the bonds in the complex involving hydrogen atoms were constrained also to their equilibrium distance [41]. The average pressure of the system was 1.0 bar. MD simulations were performed in order to monitor the dynamic behavior of the protein and the structural characteristics associated with protein–fullerene interactions, such as the number of hydrogen bonds between the ligand and the protein. A 3.5 Å distance cutoff between each donor and acceptor, along with an (donor–hydrogen–acceptor) angle cutoff of  $120^\circ$  were used as a definition for a hydrogen bonding interaction. For the trajectories obtained, further analysis (H-bond, distance and RMSD calculations) was realized with the ptraj module under AMBER.

#### MM–PBSA calculations

The binding free energy,  $\Delta G_{\text{bind}}$ , has been computed by employing the molecular mechanics Poisson–Boltzmann surface area (MM–PBSA) method [43]. This approach estimates the interaction energy in the gas phase by using MM, while the solvation free energy is determined after solving the Poisson–Boltzmann equation. The nonpolar free energy is calculated by a surface area term [44]. Computationally demanding normal mode analysis [45, 46] can be used to calculate the conformational entropy of the solute. The partition of the binding free energy into a series of

contributions (e.g. electrostatic, van der Waals etc.) allows for a valuable insight into the complex process of association [46, 47].

A brief description of the MM–PBSA methodology is given by the following equations:

$$\begin{aligned}\Delta G_{\text{bind}} &= G_{\text{complex}} - (G_{\text{HIV-1 PR}} + G_{\text{ligand}}) \\ &= \Delta H - T\Delta S \\ &= \Delta E_{\text{MM}} + \Delta G_{\text{solv}} - T\Delta S\end{aligned}\quad (1)$$

where  $\Delta H$  is the enthalpy of binding,  $T\Delta S$  is the conformational entropy contribution to  $\Delta G_{\text{bind}}$ ,  $T$  is the absolute temperature,  $\Delta E_{\text{MM}}$  is the interaction energy computed with the molecular mechanics method and  $\Delta G_{\text{solv}}$  is the solvation energy.  $\Delta E_{\text{MM}}$  and  $\Delta G_{\text{solv}}$  are given by:

$$\Delta E_{\text{MM}} = \Delta E_{\text{elec}} + \Delta E_{\text{vdW}} \quad (2)$$

$$\Delta G_{\text{solv}} = \Delta G_{\text{PB}} + \Delta G_{\text{NP}} \quad (3)$$

where  $\Delta E_{\text{elec}}$  and  $\Delta E_{\text{vdW}}$  are the electrostatic and van der Waals interaction energies, respectively;  $\Delta G_{\text{PB}}$  is the electrostatic contribution to the solvation free energy. This is computed by the Poisson-Boltzmann (PB) method [41].  $\Delta G_{\text{NP}}$  is the nonpolar solvation energy. This was computed by employing the solvent-accessible surface area (SASA) with a probe radius of 1.4 Å, according to the equation:

$$\Delta G_{\text{NP}} = \gamma \text{SASA} + \beta \quad (4)$$

for  $\gamma$ , the surface tension and the offset  $\beta$  we used the standard values of  $0.00542 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  and  $0.92 \text{ kcal mol}^{-1}$ , respectively [48]. The implementation of the method involves calculations on a series of snapshots produced by the MD; in this case, 1000 snapshots (separated by 20 ps intervals, thus spanning the whole simulation) were extracted from the 20 ns trajectory file. The values for the dielectric constant of the solvent and the solute were set to 80.0 and 1.0, respectively. The entropic contribution was calculated with the nmode module [45, 46] of AMBER, over 200 snapshots, for economy of computational time. The performance of MM–PBSA for the calculation of the binding free energy has been recently reviewed by Hou et al. [49]. The authors tested 6 different protein systems as substrates for 59 ligands. They considered various factors to study their effect to the quality of the results (e.g. the length of the MD simulation and the solute dielectric constant). Among others, the importance of the conformational entropy for the prediction of the absolute binding energy was emphasized, as well as the adequacy of MM–PBSA to reliably calculate binding free energies.

#### ADMET calculations

The QikProp program (version 3.4.109) [50] of Schrödinger molecular modeling package has been used for the

predictions of pharmacokinetic properties of fullerene derivatives **8–23** (Table 4). The program offers reliable predictions for pharmaceutically relevant properties such as cell permeability based on colorectal carcinoma (Caco-2) [51] and Madin-Darby canine kidney (MDCK) [52] cell lines. Additionally, the inhibition of hERG (human ether-a-go-go-related gene) potassium ion ( $\text{K}^+$ ) channel [53] and the central nervous system (CNS) activity [54] have been evaluated to account for the effect on toxicity of the saquinavir-based fullerene derivatives **8–23**. QikProp uses experimental results of 710 compounds, including about 500 drugs and related heterocycles. It predicts physically significant descriptors and pharmaceutically relevant properties of organic molecules, either individually or in batches. In addition to predicting molecular properties, QikProp provides ranges for comparing a particular molecule's properties with those of 95% of known drugs.

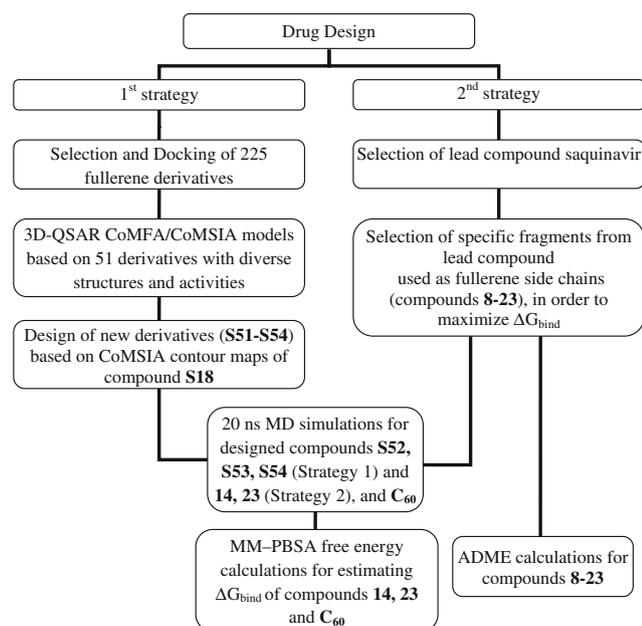
QikProp performs according to the following process. The OPLS-AA force field is used to carry out Monte Carlo (MC) simulations on organic molecules in explicit water environment. Configurational averages for a number of descriptors are next obtained, including hydrogen bond and SASA information. Correlations of these descriptors to experimentally determined properties were obtained, and then algorithms that mimic the full MC simulations and produce comparable results were developed. When evaluating a property, the program rapidly analyses atom types and charges, rotor counts, and the sample molecule's volume and surface area. The result is an accurate prediction of a molecule's pharmacologically relevant properties.

The various computational approaches we described above, along with the compounds involved in this study are schematically summarized in a flowchart (Fig. 3).

## Results and discussion

### Evaluation of docking calculations

The adequacy of the applied docking procedure was demonstrated by computing the binding energy of a series of commercially available drugs, for which this property has been determined experimentally (Table 1). It is observed that computed binding energies are in satisfactory agreement with the experimentally determined ones. The mean relative error (MRE) between experimental and calculated values is 0.19, denoting the good correlation between the values. The experimental and calculated binding energies for the commercially available drugs follow a similar trend, with the exception of amprenavir and nelfinavir. In this case, the calculated binding energy suggests that nelfinavir is a more potent HIV-1 PR inhibitor than amprenavir. A further point is that all computed



**Fig. 3** Flowchart depicting the methods employed for design and analysis

results underestimate (in absolute value) the corresponding experimental ones. The average difference between computed and experimental values is  $-2.73 \text{ kcal mol}^{-1}$ .

### 3D-QSAR CoMFA and CoMSIA validation and results

The 51 training-set compounds possess a variety of structural characteristics that affect their binding at the active site of the HIV-1 protease. A Partial Least Square (PLS) analysis using the “leave-one-out” method has been performed to cross-validate the model. The results are summarized in Table 2. The cross-validation test for the CoMFA model gives  $r_{cv}^2 = 0.613$ , while the non cross-validated value is  $r^2 = 0.977$  (Fig. S1a). The corresponding values for CoMSIA are 0.763 and 0.991 (Fig. S1b), respectively, suggesting that the performance of the latter is optimal. All values show that the proposed model is

**Table 2** A summary of the statistical results from the CoMFA/CoMSIA 3D-QSAR models

	CoMFA	CoMSIA
$r_{cv}^{2,a}$	0.613	0.763
$r^{2,a}$	0.977	0.991
No. of components	5	6
SEE <sup>b</sup>	0.154	0.137
$F_{test}^c$	128.614	278.614
Contributions		
Steric	0.729	0.115
Electrostatic	0.271	0.011
Hydrophobic	N/A	0.374
Acceptor	N/A	0.257
Donor	N/A	0.243

<sup>a</sup>  $r_{cv}^2$  and  $r^2$  are the cross-validated (cv) and non-cross validated terms, respectively

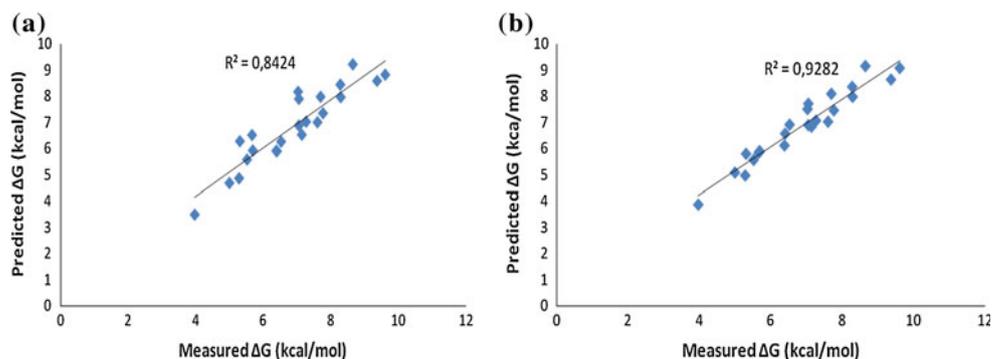
<sup>b</sup> SEE standard error of estimate

<sup>c</sup>  $F_{test}$  = ratio of explained to unexplained values

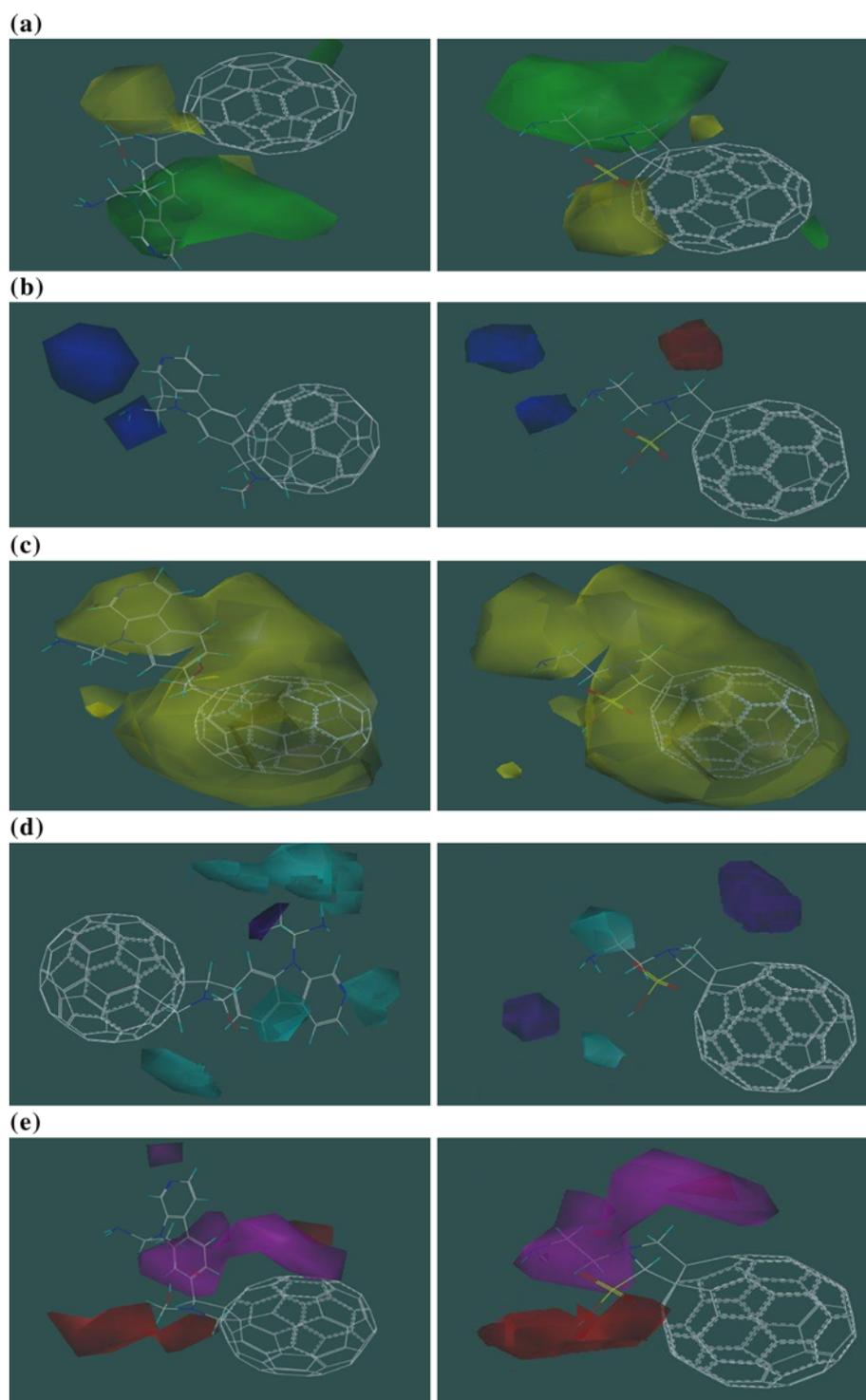
statistically significant. The CoMFA model shows that the relative steric and electrostatic contributions are 72.9% and 27.1%, respectively. Steric, electrostatic, hydrophobic, hydrogen donor and hydrogen acceptor contributions for CoMSIA are 11.5%, 1.1%, 37.4%, 25.7% and 24.3%, respectively. Both CoMSIA and CoMFA show that the steric interactions are considerably larger than the electrostatic ones (Table 2). According to the CoMSIA model, the dominant contribution is associated with hydrophobic interactions. These results agree with the observations that the fullerene core is mostly hydrophobic [12].

The performance of the developed QSAR models has been validated with a test set that includes experimentally determined binding energies [55–57] (Table S2). The experimental binding energies show good correlation with the CoMFA/CoMSIA predicted values ( $R_{CoMFA}^2 = 0.84$ ,  $R_{CoMSIA}^2 = 0.92$ , Fig. 4a, b, respectively). It is observed that the CoMFA/CoMSIA results are in agreement with the experimental values, however the performance of CoMSIA

**Fig. 4** Graphs of experimental and predicted binding energies ( $\text{kcal mol}^{-1}$ ) of Table S2 from **a** CoMFA and **b** CoMSIA



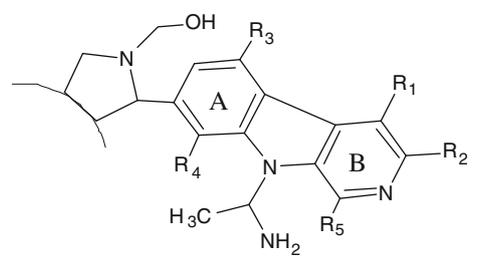
**Fig. 5** High binding energy compound **S18** (left), and low binding energy compound **S4** (right) contour maps produced from the CoMSIA 3D-QSAR model that show the relative contributions: **a** green (yellow) color shows the sterically favored (disfavored) area. **b** electrostatic interactions in blue (red) denote regions where electronegative groups increase (decrease) the affinity of the ligand. **c** hydrophobic interactions in yellow (white) denote regions where hydrophobic groups increase (decrease) the inhibitory effect. **d** H-bond donor contour map in cyan (purple) shows the favored (disfavored) areas. **e** H-bond acceptor contour map in magenta (red) color shows the favored (disfavored) areas

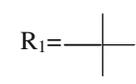
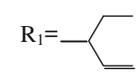
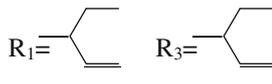
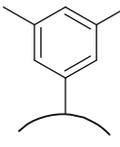
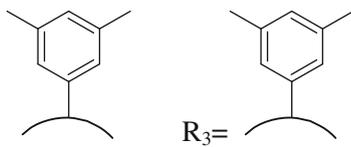
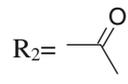


appears better. Compared to a previous work from Durdagi et al. [17] the model proposed in this article has a better cross-validated  $r^2$  value ( $r_{cv}^2 = 0.763$  as opposed to  $r_{cv}^2 = 0.670$ ).

CoMSIA electrostatic contour maps are presented for compound **S18** (Table S1; Fig. 5, left), which has one of the highest binding energies of the training set. It is observed that there is a sterically disfavored area close to

the fullerene core (Fig. 5a, left), while there is an area near the aromatic rings, which favors steric interactions. The hydroxyl ( $-\text{OH}$ ) and amino ( $-\text{NH}_2$ ) groups can form hydrogen bonds (Fig. 5d, e) and they also participate in electrostatic interactions (Fig. 5b). Certain parts of aromatic rings A and B (Table 3) act as H-bond donors, while ring A appears also to be electron acceptor (Fig. 5d, e).

**Table 3** Calculated binding energies of modified compound **S18** (see Table S1) based on the CoMSIA contour maps


Comp.	Substituents	Calculated binding energy (kcal mol <sup>-1</sup> )
<b>S18</b>	R <sub>1</sub> = R <sub>2</sub> = R <sub>3</sub> = R <sub>4</sub> = R <sub>5</sub> = -H	-10.82
1		-13.32
2		-14.02
3		-13.25
4		-13.45
5		-12.03
6	R <sub>2</sub> = -COOH	-14.08
7		-12.68

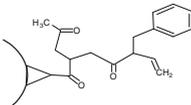
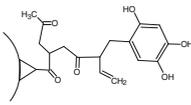
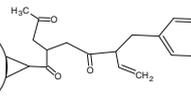
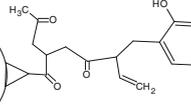
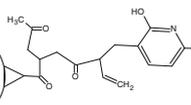
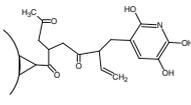
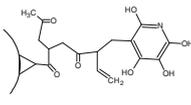
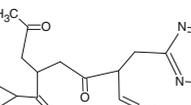
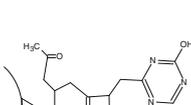
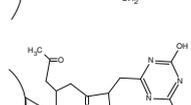
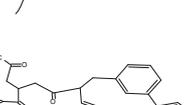
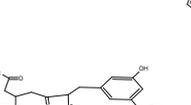
CoMSIA electrostatic contour maps for compound **S4** (Table S1; Fig. 5, right) are also presented. This compound has one of the lowest binding energies. The steric component (Fig. 5a, right) has the favored (green) areas close to the N-containing ring and its side chain, while the area between the fullerene cage and the -SO<sub>2</sub> group is disfavored (yellow). The region, which favors electrostatic interactions (Fig. 5b, right, blue) does not overlap with **S4** and that could be a reason for the reduced activity of this compound. Figures 5d and e (right) show the H-bond contributions (donor and acceptor, respectively). In Fig. 5 the magenta color shows the favored contributions that stem from the terminal -NH<sub>2</sub> group, while the -SO<sub>2</sub> group (red) represents the less favored contributions. The contour

maps have been generated with respect to derivatives **S18** and **S4**, with high and low binding affinities, respectively. The point of importance here is the discussion on the useful information (regarding optimal binding), which can be extracted from the contour maps.

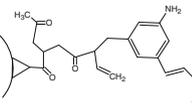
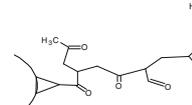
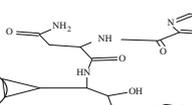
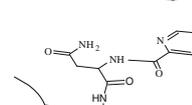
#### First strategy for the design of HIV-1 PR Inhibitors

We have used the CoMSIA contour maps of **S18**, in order to design compounds with optimized binding affinity to HIV-1 PR. The proposed compounds are given in Table 3. Addition of steric groups such as tertiary butyl (1), pentane (2) and 1,3-dimethyl benzene (4) shows an increase (in absolute value) of the binding energy, as depicted by the values of

**Table 4** Proposed structures and their calculated binding energies

No.	Compound	Binding energy (kcal mol <sup>-1</sup> )
8		-13.44
9		-14.34
10		-14.30
11		-14.05
12		-14.33
13		-15.03
14		-15.93
15		-14.48
16		-14.10
17		-14.57
18		-13.46
19		-14.35

**Table 4** continued

No.	Compound	Binding energy (kcal mol <sup>-1</sup> )
20		-14.40
21		-14.23
22		-15.83
23		-16.03

-13.32 kcal mol<sup>-1</sup>, -14.02 kcal mol<sup>-1</sup>, and -13.45 kcal mol<sup>-1</sup>, respectively. When a second steric group (compounds 3/5) is added, the binding energy decreases compared to the mono-substituted derivatives (-13.25 kcal mol<sup>-1</sup> and -12.03 kcal mol<sup>-1</sup>, respectively). This decrease may be due to the bulkier nature of the fullerene analogues and the steric hindrances that may arise within the binding site. Addition of electronegative groups, such as -COOH (6) has also resulted in increase of the binding energy (-14.08 kcal mol<sup>-1</sup>; Table 3).

#### Second strategy for the design of HIV-1 PR inhibitors

This strategy involves: (1) The choice of a compound, which is known to have good binding affinity for HIV-1 PR. For this purpose, we have selected the drug saquinavir (Table 1), (2) The choice of a functional group of saquinavir, which will be appropriately modified and attached to the fullerene core, in order to produce a compound with improved binding affinity. The selection of the groups is shown in Fig. S2 in supplementary material. The results for the modified fullerenes are given in Table 4. It is noted that compounds 11–14 show the effect on the binding energy, after substituting 1 to 4 -H of an aromatic ring with -OH groups. In fact, the maximal favorable effect is achieved by replacing all four hydrogen atoms (Compound 14). A considerable increase (in absolute values) of the binding energy is achieved by replacing -C<sub>6</sub>H<sub>5</sub> by -C<sub>6</sub>H<sub>4</sub>N as shown in compounds 8 and 10. Also comparison of compounds 9 and 13 shows that the greater positive effect

on the binding energy is achieved when the  $-C_6HN(OH)_3$  group is present rather than  $-C_5H_2(OH)_3$ . On the other hand, compounds **10/11** and **15/16** show a small decrease in binding energy following the substitution of a  $-H$  with a hydroxyl group. Substitution of three  $-H$  by  $-OH$  (compound **19**) or  $-NH_2$  groups (compound **20**) leads to an increase of the binding affinity. Finally, for the design of derivatives **22** and **23**, two fragments of saquinavir, which possess peptide-like characteristics, were employed. The presence of the amide bond apparently increases the possibility of H-bonding and thus it can explain the larger binding energies associated with these compounds.

Particularly potent inhibitors appear to be compounds **13**, **14**, **21**, **22** and **23**. The last two designed inhibitors involve amide bonds in the side chain of the fullerene. To the best of our knowledge, the best binding affinity of a fullerene derivative toward HIV-1 PR so far is 36 nM [18]. The compounds we have designed (Tables 3, 4) presented binding affinities in the range of pM, for example compound **14** has an affinity of 28.3 pM, which is better than that of lopinavir. Since, on average, the docking procedure underestimates the binding energy by approximately  $-2.73 \text{ kcal mol}^{-1}$  (Table 1), one may estimate that the binding energy of compound **14** is around  $-18.66 \text{ kcal mol}^{-1}$  and that of compound **23** is around  $-18.76 \text{ kcal mol}^{-1}$ .

#### Hydrogen bonding analysis

The hydrogen bonding analysis of fullerene–HIV-1 PR complexes reveals differences in binding modes. The H-bonds between the fullerene analogues (**S52**, **S53** and **S54** associated with the lowest binding energies and **14**, **23**

associated with the highest binding energies) and the protease involve one or more of the following residues: the aspartic acids in positions 29/29', 30/30' and the glycines in flap positions 48/48' (see Table 5 for specific interactions). All major H-bond patterns refer to interactions between backbone amide hydrogen atoms (N–H) of the protease and fullerene oxygen atoms (OH, C–O–C or C=O). The results of Table 5 are expressed as occupancy percentage of H-bonds between fullerene analogues and HIV-1 PR. For example, a hydrogen bond between Asp29 and compound **S52** appears for 61% of the simulation time, whereas a strong interaction between compound **14** and Asp29' exists throughout the simulation (91%). Also note that within a particular complex, the fullerene prefers to form H-bonds with residues in one of the two chains of the protease. In compound **S52** an important fullerene-flap interaction occurs mainly during the first 5 ns between the carbonyl O1 of the fullerene and Gly48 of the protease (Table 5). An interesting observation is that H-bond interactions involving compound **S54** are determined by the two equivalent oxygen atoms on the fullerene side chain: during the first half of the simulation, the structure of the complex is mainly stabilized by Asp29 and Asp30 interactions with O1 and O2, respectively, whereas O1 is primarily associated with Asp30 (and O2 with Asp29) for the last 5 ns (Fig. S2c (i) and S2c (ii)). Another important observation is that compound **14** has more groups participating in H-bonds than the other derivatives (Table 5; Fig. 6); a H-bond (O7 Fig. S3d) with one flap (Gly48') and another (O4) with Asp29' are the principal interactions that render **14** as one of our most efficiently bound structures.

The analysis performed on the complex of the protease with compound **23** revealed only two hydrogen bonds with

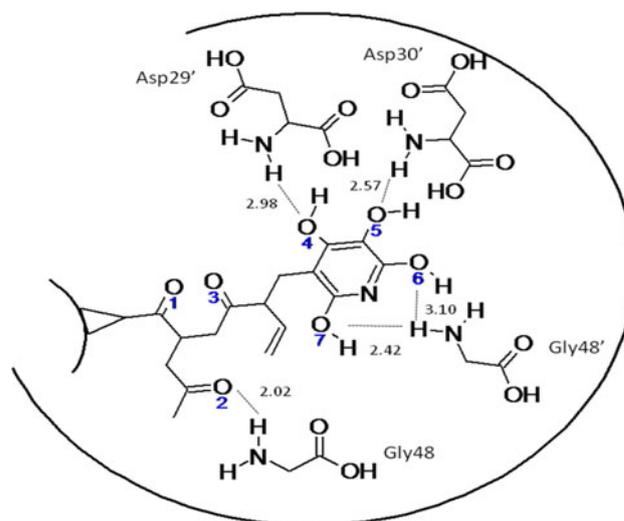
**Table 5** Occupancy percentage of H-bonds between fullerene analogues and HIV-1 PR residues as calculated from the MD simulations

Compounds <sup>a</sup>					
Residues <sup>b</sup>	S52	S53	S54	14	23
ASP29	60.57	N/A <sup>c</sup>	75.31	N/A	N/A
ASP30	22.14	N/A	12.78	N/A	N/A
ASP29'	N/A	78.42	N/A	93.51	N/A
ASP30'	N/A	39.97	N/A	47.71	N/A
GLY48	45.03	N/A	N/A	13.85	25.70
GLY48'	N/A	N/A	N/A	60.24	N/A
ASP25'	N/A	N/A	N/A	N/A	10.00
Calculated binding energy (kcal mol <sup>-1</sup> )	8.28	6.38	5.52	15.93	16.03

<sup>a</sup> Compounds used in MD calculations (Tables 3, 4)

<sup>b</sup> Amide groups (–NH) from residues of the protein participating in H-bonds with the fullerene analogues

<sup>c</sup> N/A: H-bonds occurring less than 10% of the simulation time are not shown



**Fig. 6** Schematic representation of hydrogen bonds between compound **14** and residues in the binding pocket of the protease

the protein. Similar to other derivatives analyzed, compound **23** formed one bond with Gly48 at the flap of the protease (present for 26% of the simulation time) and a less frequent bond (10%) with Asp25' at the active site of the protein. This particular derivative is the only one that showed formation of a bond with catalytic residues. Thus, we observed that even though the selected fullerenes (compounds **S52**, **S53**, **S54**, **14**, **23**) do not interact with the same residues, all of them form H-bonds with one or more residues near the active site (29, 30, 29' and 30'), as well as with flap residues 48 and 48'. The H-bond analysis on compounds **S52**, **S53**, **S54** and **14** suggests a correlation between the number of H-bonds and binding energy (Table 5). However, compound **23** does not conform to this trend. This indicates that there are other factors as well, strongly influencing the binding energy. This will be discussed in the MM–PBSA analysis section.

#### Analysis of active site–flap and flap–flap distances

We have calculated the distances between the C $\alpha$  atoms of the catalytic Asp25 and flap tip Ile50, as well as between the C $\alpha$  atoms of flap tips Ile50 and Ile50', for the complex of the protease with compound **14**, which has one of the highest binding energies of the proposed derivatives. The results are depicted in Fig. 7. We observed a significant change in the distance between the two isoleucines (Ile50 and Ile50') (Fig. 7b). During the first 10 ns of the simulation, the distance between the flaps of the protease increases and reaches a plateau at  $\sim 10$  Å; afterwards it remains constant for the rest of the simulation. This suggests the stabilization of the flap structure on top of the ligand, resulting in a closed, bound–HIV-1 PR conformation. A similar feature has been also observed by Zhu et al. [31]. The distance between Asp25 and Ile50 is greater in the first nanoseconds (Fig. 7a), slowly decreases and eventually stabilizes around 17.5 Å. According to other experimental and theoretical studies, this indicates a shift of the flaps towards the active site of the protease, thus explaining the typical behavior of HIV-1 PR upon binding [6, 58]. During the simulation, at around 7 ns, 13 ns

and 17 ns the distance drops to its lowest value ( $\sim 15.5$  Å). The small fluctuation of the distance between the tips of the flaps after the 10 ns of the simulation can be explained by the presence of an H-bond between Gly51 and Gly51' (65% occupancy). This bond appears at approximately 7.5 ns and is present for the remaining of the simulation, accounting for the earlier observation. A hydrogen bond also appears between the two Ile residues (22% occupancy) but not with the same frequency as that of the Gly51/51'.

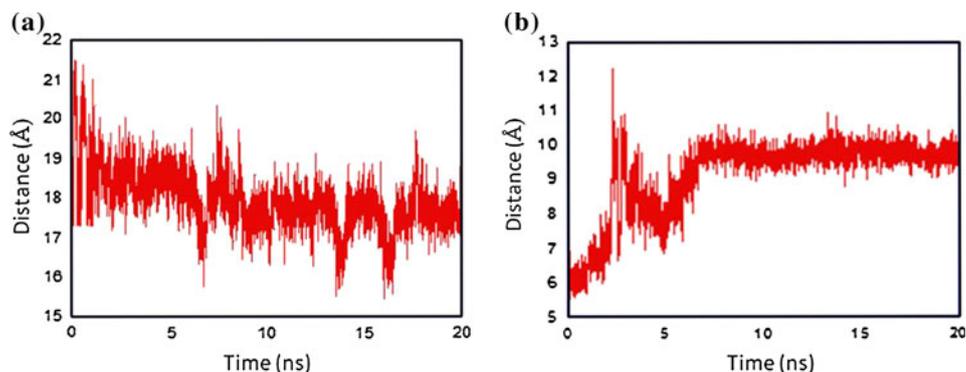
Our results show (see also hydrogen bonding analysis) that the fullerene analogues do not bind directly to the triplet of Asp–Gly–Thr, they rather interact mostly with other residues near the active site (29, 29', 30, 30') and with the flaps of the protease (e.g. Gly48/48'). The complex of HIV-1 PR with compound **23** also follows similar trends with compound **14** regarding the variation of distances. The only difference is focused on the interaction between Gly51 and Gly51', which is less dominant in the former case (38% occupancy). The different ligands studied here seem to interact with two different regions (Fig. 1, active site–red, flap region–blue) of the protease. These interactions may account for the strong binding affinity of compounds **14** and **23** to HIV-1 PR.

#### RMSD analysis

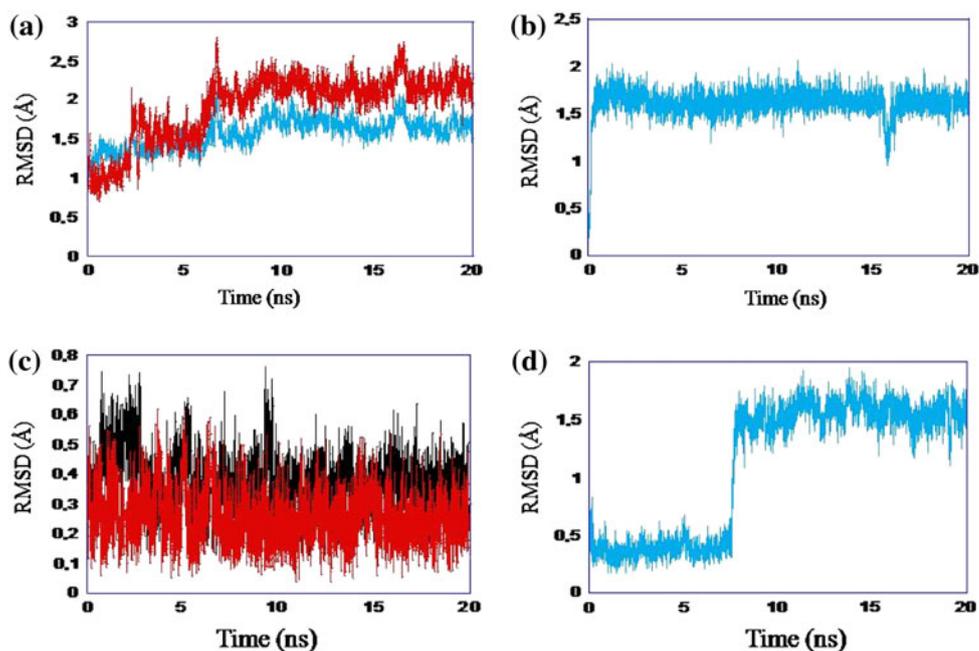
To further test the stability of the protease, we performed RMSD analysis for C $\alpha$  atoms of the **14**–HIV-1 PR complex, using as reference structure the one obtained after docking. This analysis has shown that there is no major structural change in the protein during the simulation; the average value for HIV-1 PR is 1.9 Å (Fig. 8a, blue). The RMSD values converge after approximately 10 ns.

RMSD analysis has been also performed for the flap region of the complex, with selected residues 44–55 and 44'–55' (Fig. 8a, red). The results clearly show the structural change in the flap region (average RMSD = 2.6 Å). The initially unstable structure of the flap region eventually converges to a less mobile state, mainly because of the stabilizing hydrogen bonding interactions between the

**Fig. 7** Calculated distances over 20 ns between C $\alpha$  atoms for the residues Asp25 (active site), Ile50 and Ile50' (flap tip) for the complex of HIV-1 PR with compound **14**: **a** Asp25–Ile50 distance. **b** Ile50–Ile50' distance



**Fig. 8** RMSD distribution over time, for HIV-1 PR complexed with compound **14** (a, b, c) and with compound **23** (d). **a** All protease residues (blue) and residues (44–55, 44'–55') in the flap region (red). **b** and **d** Active site residues (Asp-Thr-Gly). **c** Residues 79–83 (red) and 79'–83' (black)



ligand and Gly48' at the flaps of the protein (Table 5) and between the two Gly residues at the flap region. We observe that even though the deviations in the flap structure are greater than these of the whole protein, they follow a strikingly similar pattern. This allows for the conclusion that RMS deviations observed for HIV-1 PR are mainly due to the structural changes of the flaps, with the strong suggestion that the rest of the protease remains relatively stable during the simulation.

The following step was to look at the RMSD for the residues at the active site of the protease (Fig. 8b). The RMSD ( $\sim 1.5$  Å) shows that active site residues after an initial structural change do not deviate significantly for the rest of the simulation. Subsequently, we performed an RMSD analysis on a region surrounding the active site. The area defined by residues 79–83 (and 79'–83') covers the active site and comes in contact with the solvent (Fig. 1, purple). We observed (Fig. 8c) that there is no large difference between these two regions of the protease. The trend in RMSD for this regions shows similarities with previous observations in other HIV1-PR–ligand complexes studied by Zoete et al. [59] and further confirms a well-defined bound structure for HIV-1 PR.

RMSD analysis for the complex of compound **23** with HIV-1 PR has shown that there are similarities with the complex of compound **14**. The flap region of the protease shows again larger variation in RMSD values in comparison with the whole protein (Fig. S4). A difference is observed at the active site region where the RMSD remains

very low ( $\sim 0.5$  Å) during the first 7 ns, and after that time it reaches the same average value (1.5 Å, Fig. 8d) as **14**. The interaction between Asp25 and compound **23** may have induced the structural change of the active site during the simulation.

It is evident from the RMSD analysis that there are no significant structural changes in the protease (other than in the flap region), following ligand binding in the active site cavity. A stabilization, as expected, for the flap region of the protein is eventually observed during the simulation, due to the presence of interactions between the flaps and the ligand (H-bonds with Gly48/48').

#### MM–PBSA analysis

To evaluate the energetics of binding in a more reliable and detailed way, we have employed the MM–PBSA method to specific fullerene analogues. The convergence of the procedure has been achieved for HIV-1 PR complexes with compounds **14**, **23**, and  $C_{60}$  after approximately 7 ns, as depicted in Fig. S5 ( $\Delta H$  vs. time) and Fig. S6 ( $\Delta E_{vdW}$  vs. time). The MM–PBSA results of compounds **14** and **23** are given in Table 6. For comparison, calculated  $\Delta G_{bind}$  of  $C_{60}$  is also presented. It is observed that both docking and MM–PBSA results for compounds **14** and **23** agree qualitatively. Both methods (1) show that  $|\Delta G_{bind}|$  of compound **23** is slightly larger than the corresponding value for compound **14** and (2) confirm that compounds **14** and **23** have large  $|\Delta G_{bind}|$ . Compounds **14** and **23** have very similar  $\Delta E_{MM}$

**Table 6** Contributions to  $\Delta G_{\text{bind}}$  for compounds **14**, **23** and  $C_{60}$  complexed with HIV-1 PR, computed with the MM–PBSA method

Energetic analysis	$\Delta G_{\text{bind}}$ (kcal mol <sup>-1</sup> )		
	Comp. <b>14</b>	Comp. <b>23</b>	$C_{60}$
$\Delta E_{\text{elec}}$	-14.74 (3.01) <sup>a</sup>	-20.78 (2.70)	-5.24 (3.14)
$\Delta E_{\text{vdW}}$	-83.90 (2.55)	-76.80 (2.61)	-65.37 (3.00)
$\Delta E_{\text{MM}}$	-98.65 (2.48)	-97.58 (2.94)	-70.61 (2.89)
$\Delta G_{\text{NP}}$	-26.35 (2.11)	-25.14 (2.16)	-34.25 (2.95)
$\Delta G_{\text{PB}}$	59.92 (2.32)	60.37 (1.99)	56.59 (2.58)
$\Delta G_{\text{solv}}$	33.57 (2.13)	35.23 (2.06)	22.34 (3.04)
$\Delta G_{\text{elec(tot)}}$	45.18 (2.44)	39.59 (2.02)	51.35 (2.76)
$\Delta G_{(\text{MM}+\text{solv})}$	-65.08 (2.87)	-62.35 (2.30)	-48.27 (2.92)
$-T\Delta S_{\text{tot}}$	53.11 (2.24)	49.60 (1.95)	49.52 (3.14)
$\Delta G_{\text{bind}}$	-11.97	-12.75	-1.25

Note  $\Delta G_{\text{elec(tot)}} = \Delta E_{\text{elec}} + \Delta G_{\text{PB}}$

<sup>a</sup> Numbers in parentheses denote standard deviations

values, although their  $\Delta E_{\text{elec}}$  and  $\Delta E_{\text{vdW}}$  values are different. Most of the other contributions to  $\Delta G_{\text{bind}}$  are rather similar in both complexes. It is also observed that  $\Delta G_{\text{PB}}$  is much larger than  $|\Delta E_{\text{elec}}|$ . The total electrostatic contribution is 45.18 and 39.59 kcal mol<sup>-1</sup>, for compounds **14** and **23**, respectively. Thus, the unfavorable electrostatic contribution associated with solvation is not fully compensated by the favorable contribution to the MM energy. Additionally, the significant value of the entropic term ( $-T\Delta S$ ) for all three derivatives is noted; in fact, all compounds have comparable values for  $-T\Delta S$ . The binding of compounds **14**, **23** and  $C_{60}$  into HIV-1 PR is mainly driven by  $\Delta E_{\text{vdW}}$  (which makes the single largest contribution in absolute value) and  $\Delta G_{\text{NP}}$ . This trend has been verified by several studies previously [59–61].

Concerning  $C_{60}$ , we observe that it has a small  $|\Delta G_{\text{bind}}|$ . The large  $|\Delta G_{\text{bind}}|$  of compounds **14** and **23** is thus associated with the corresponding substituent. The presence of

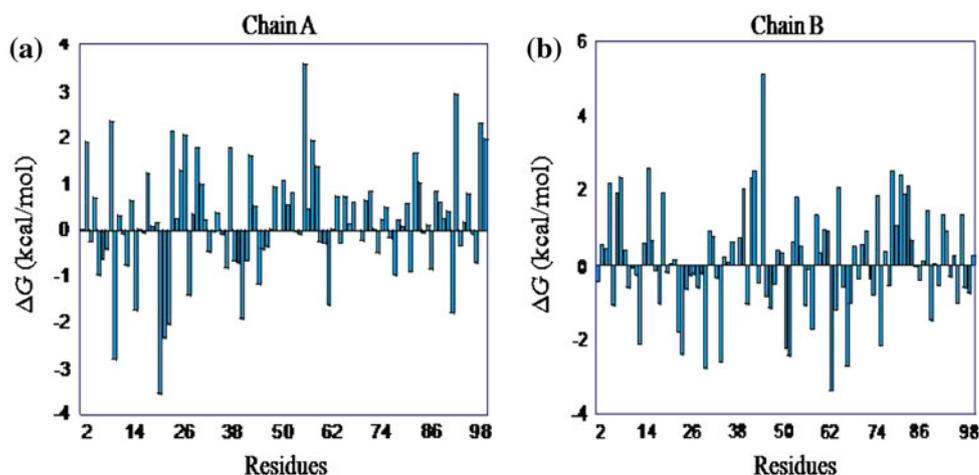
$C_{60}$  core in the studied derivatives is primarily associated with the high contribution of van der Waals interactions ( $\Delta E_{\text{vdW}}$ ) to  $\Delta G_{\text{bind}}$ .

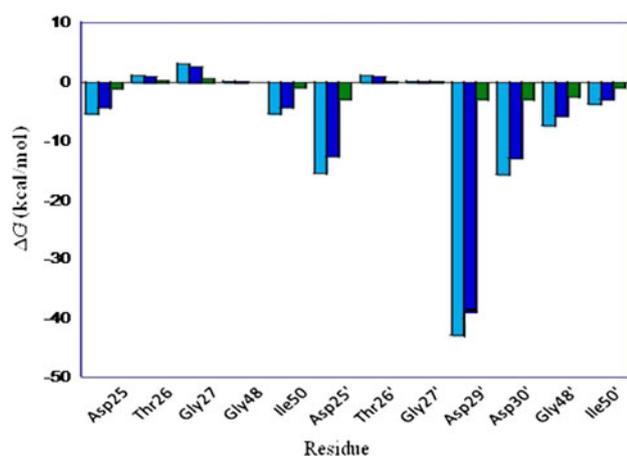
Instead of modeling the water around a protein by explicitly simulating the motions of thousands of water molecules, MM–PBSA discards the explicit water molecules and uses a parametrized implicit water model (PB) consisting of new energy terms for the hydrophobic effect. This is an obvious drawback that may have accounted for the difference between our calculations and the experiment. Apart from the aforementioned limitation however, the computationally efficient LCPO method for estimating the hydrophobic contribution (SASA) in MM–PBSA appears to perform relatively well even for hydrophobic systems [48, 62, 63], reproducing the experimental results in a satisfactory way [64].

### Free energy decomposition

The next step was to analyze the different components of the binding energy, as well as the contribution of the protease residues in the binding free energy. Figure 9 shows a representation of the energy contribution ( $\Delta G_{\text{bind}}$ ) for each amino acid in the protease, in both chains. Energy decomposition showed that certain residues at the flaps of the protease (in only chain B) have favorable contributions to the total free binding energy; for instance, residues Gly48' and Gly51' have contributions of  $-2.7$  and  $-2.4$  kcal mol<sup>-1</sup>, respectively. We notice that the enhanced H-bond interactions between compound **14** and the flap of chain B are accompanied with a significant contribution to  $\Delta G_{\text{bind}}$ , in contrast to flap residues in chain A. Additionally, favorable contribution to the binding energy comes also from residues Glu21 and Ile62' (Fig. 9,  $-3.3$  kcal mol<sup>-1</sup> for both residues). Finally, it is observed

**Fig. 9** Per residue contribution to free energy of binding  $\Delta G_{\text{bind}}$  for (a) chain A, and (b) chain B of the protease for the complex with compound **14**





**Fig. 10** Per residue contributions  $\Delta H$  (cyan),  $T\Delta S$  (blue),  $\Delta G$  (green) for residues of the active site (Asp-Thr-Gly) and flap region (Ile50/50'), as well as residues involved in H-bonding with compound **14** (Asp29', Asp30', Gly48, Gly48')

that residues around the active site (in both chains) contribute significantly to the binding energy.

Figure 10 shows the energy contribution ( $\Delta H$ -cyan,  $T\Delta S$ -blue,  $\Delta G_{\text{bind}}$ -green) for residues of the active site (Asp25/25'-Thr26/26'-Gly27/27') and flap tips (Ile50/50'), as well as residues involved in H-bonding with compound **14** (Asp29', Asp30', Gly48, Gly48'). Asp25' has the highest contribution to the binding energy ( $-2.7 \text{ kcal mol}^{-1}$ ) from all residues in the active site of the protease, even though it is not associated with compound **14** via a H-bond (Fig. 10). The permanent H-bond between the ligand and Asp29' may have accounted for the largest contribution to the total binding energy, as depicted in Fig. 10. It can be concluded

that although the presence of an H-bond does not guarantee the favorable energy contribution of the residue involved, residues that contribute favorably are usually involved in H-bonding with the ligand.

A general trend observed in Fig. 10 is that the enthalpic ( $\Delta H$ -cyan) contribution is higher than that of the entropic contribution ( $\Delta S$ -blue). However, their values are comparable. It is evident from the aforementioned results (Fig. 9) that several residues, not directly interacting with the ligand, contribute favorably to the binding energy of the complex. In conclusion, most of the residues that have high energy contributions are close to the active site and flap regions of the protease, suggesting that these two areas of the enzyme play an important role in the binding of ligands.

#### ADMET predictions for saquinavir-related fullerene derivatives

ADMET properties have been calculated for the proposed fullerene derivatives **8–23** (Table 7). Although usually fullerene-derived HIV-1 PR inhibitors are known to have relatively low toxicity [9], if specific functional groups (i.e. saquinavir and its derivatives) were added they may result in problematic drug candidates. Especially, since saquinavir is known to have significant side effects [65], pharmacokinetic profiles of the new compounds have to be estimated. For this aim, we used QikProp to predict the following pharmacokinetic properties: cell permeability based on Caco-2 and MDCK models, inhibition of hERG  $K^+$  channel and CNS activity.

**Table 7** Predicted pharmacokinetic profiles of proposed saquinavir-modified fullerene derivatives

Compound No.	CNS activity (– – to ++)	hERG K + Channel blockage (log $IC_{50}$ )	Apparent Caco-2 permeability (nm/s) (<25 poor, >500 great)	Apparent MDCK permeability (nm/s) (<25 poor, >500 great)
8	–	–7.746	1365	692
9	– –	–7.323	65	26
10	– –	–7.389	736	355
11	– –	–7.372	341	155
12	– –	–7.271	104	43
13	– –	–7.068	40	15
14	– –	–6.977	22	8
15	– –	–7.195	423	195
16	– –	–7.090	133	55
17	– –	–7.006	43	16
18	– –	–9.501	1447	737
19	– –	–9.064	39	14
20	– –	–9.220	102	42
21	– –	–7.587	1482	756
22	– –	–6.606	9	14
23	– –	–6.848	14	19

The Caco-2 and MDCK cells are widely used to predict the absorption of drug candidates across cell barriers. For both cell types, an ideal permeability is expressed above 500 nm/s, whereas rates lower than 25 nm/s are considered to be poor [50, 66]. Thus, most of the proposed fullerene derivatives (compounds **10**, **11**, **12**, **15**, **16**, **20** and especially compounds **8**, **18**, **21**) have good permeability at both cell types. It is worth noting that compounds with polar groups (such as –OH in compounds **9**, **13**, **14**, **19**) present low permeability values. In contrast, compounds with reduced polar character showed an increased absorption potential. Another important observation indicated the influence of the structure of saquinavir on cell permeability. We observed that compounds **22** and **23** have very low permeability values in both cells (Table 7) compared to other compounds. Even though all side chains of fullerene compounds **8–23** are saquinavir fragments, larger saquinavir moieties have been attached to compounds **22** and **23** (see Fig. S2 and Table 4). The increased saquinavir-like character of compounds **22** and **23** may have additionally contributed to the low Caco-2 and MDCK permeability. This is not surprising since it is known that saquinavir has low intestinal permeability; indeed, clinical studies have measured saquinavir permeability to Caco-2 and MDCK to be  $9.3 \pm 1$  nm/s and  $4.63 \pm 0.25$  nm/s, respectively [67]. In conclusion, it was observed that increased polar character along with great resemblance to saquinavir structure lowered the permeability values for the drug-modified fullerene compounds. Nevertheless, polar groups are usually needed because they induce favorable interactions between the drug and cavity residues. Therefore, successful drug design should aim at maximizing the polar character of an inhibitor to the point where its cell permeability remains adequate.

Additionally, QikProp predicted CNS activity for compounds **8–23** on a –2 (inactive) to +2 (active) scale. From the tested molecules, **8** and **21** are more CNS active compared to others however, it was obvious that all compounds presented very low activity, thus avoiding undesirable CNS side effects [20].

The hERG K<sup>+</sup> channel has been implicated as a cardiotoxicity marker. Therefore, hERG potassium channel screening for ADME and toxicity properties of potential hit compounds is exceedingly important and is gaining acceptance in the pharmaceutical industry. Investigations have revealed that blockage of hERG channels lead to LQT syndrome [55]. Proposed compounds (except **18–20**), have more than nM IC<sub>50</sub> at the central cavity of the channel. This indicates a relatively low toxicity for the proposed derivatives. Additionally, compounds **14**, **22** and **23** were predicted to have the lowest toxicity among all others; as mentioned above, their greater resemblance to saquinavir compared to the other compounds may have also attributed to the reduced toxicity of compounds **14**, **22** and **23**

(experimental hERG K<sup>+</sup> channel blockage values for saquinavir: IC<sub>50</sub> = 15.3 μM, log IC<sub>50</sub> = –4.82) [68].

## Conclusions

The objectives of this study were (1) To use efficient strategies for the design of novel fullerene-based inhibitors for HIV-1 PR. These strategies are applicable to design drugs for any other class of proteins, and (2) To understand the molecular mechanism associated with the binding pattern of the proposed inhibitors.

First, we confirmed the adequacy of the docking procedure we used. Second, we developed two 3D-QSAR models, CoMFA and CoMSIA. The training set involved 51 molecules with quite varying activity. Reliable statistical parameters were obtained for both models. Their predicting ability has been demonstrated by using a set of experimental results. CoMSIA shows that the highest contributions to the binding energy are associated with hydrophobic interactions and H-bonding. Employing our CoMSIA model, we designed a series of novel fullerene-based inhibitors. A second strategy for designing inhibitors was based on appropriately modified fragments of the drug saquinavir. Most of the designed fullerene derivatives had very good binding energies with HIV-1 PR.

We have investigated the H-bond patterns between compounds **14**, **23** (they presented the highest binding energies) and HIV-1 PR. We observed that even though the selected fullerenes do not form hydrogen bonds with the same residues, all of them: (1) form H-bonds with residues near the active site (29, 30, 29' and 30'), or with residues lying at the flap (48 and 48') region, and (2) do not bind directly to the active-site triplet of Asp-Gly-Thr.

We found that the flaps of the protease show structural changes during the simulation, a trend also observed in previous experimental and theoretical studies. A stabilization of the flap region of the protein has been observed during the simulation. This stabilization has been associated with interactions between the flaps and each fullerene derivative (e.g. H-bonds with Gly48/48').

Our MM-PBSA results showed that the binding of the considered derivatives (**14**, **23** and C<sub>60</sub>) to HIV-1 PR was mainly driven by van der Waals and nonpolar contributions ( $\Delta E_{vdW}$  and  $\Delta G_{NP}$ ). Analysis of the results has shown that the large  $|\Delta G_{bind}|$  of compounds **14** and **23** is due to the selected substituents. The large contribution of  $\Delta E_{vdW}$  to  $\Delta G_{bind}$  is associated with the C<sub>60</sub> core.

Free energy decomposition indicated that the enthalpic contribution is higher than that of the entropic one, even though they are comparable. Of particular importance for the binding were areas close to the active site and the flap regions of the protease. On the other hand, it was observed

that binding also depends on a set of residues, not directly involved in H-bond interactions.

Finally, estimation of ADMET properties for saquinavir-related fullerene derivatives **8–23** revealed that most of the designed compounds have adequate Caco-2 and MDCK permeability values, they are primarily CNS-inactive and they presented a limited ability to block hERG K<sup>+</sup> channel, thus indicating a relatively low toxicity. Furthermore, compounds **14**, **22** and **23** which have the greatest structural similarity to saquinavir appeared the least toxic. A final observation related the increased number of polar groups within a compound and its increased saquinavir-like structure with a reduced cell permeability.

**Acknowledgments** We acknowledge the CINECA award under the IS CRA initiative, for the availability of high performance computing resources and support. This work was supported by funding provided by the European Commission for the FP7-REGPOT-2009-1 Project ‘ARCADE’ (Grant Agreement No. 245866). Also, the research has been co-financed by the European Union (European Social Fund-ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF)–Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.

## References

- Source available at <http://www.unaids.org/>
- Hamacher K (2008) *Gene* 422:30
- Navia MA, Fitzgerald PMD, McKeever BM, Leu CT, Heimbach JC, Herber WK, Sigal IS, Darke PL, Springer JP (1989) *Nature* 337:615
- Jewell NA, Chen R, Raices R, Mansky LM (2003) *J Antimicrob Chemother* 52:547
- Trylska J, Tozzini V, Chang CA, McCammon JA (2007) *Bioph J* 92(12):4179
- Hornak V, Okur A, Rizzo RC, Simmerling CJ (2006) *Am Chem Soc* 128:2812
- Wensing AMJ, van Maarseveen NM, Nijhuis M (2010) *Antiviral Res* 85:59
- Hansch C, Unger SH (1977) *J Med Chem* 16(11):1217
- Friedman HS, DeCamp DL, Sijbesma RP, Srdanov G, Wudl F, Kenyon G (1993) *J Am Chem Soc* 115:6506
- Toniolo C, Bianco A, Maggini M, Scorrano G, Prate M, Marastoni M, Tomatis R, Spisani S, Palu G, Blair ED (1994) *J Med Chem* 37:4558
- Jafvert CT, Kulkarni PP (2008) *Environ Sci Technol* 42:5945
- Marchesan S, Da Ros T, Spalluto G, Balzarini J, Prato M (2005) *Bioorg Med Chem Lett* 15:3615
- Zhang HW, Coats SJ, Bondada L, Amblard F, Detorio M, Asif G, Fromentin E, Solomon S, Obikhod A, Whitaker T, Sluis-Cremer N, Mellors JW, Schinazi RF (2010) *Bioorg Med Chem Lett* 20:60
- Lee VS, Nimmanpipug P, Aruksakunwong O, Promsri S, Sompornpisut P, Hannongbua S (2007) *J Mol Graph Model* 26:558
- Troshina OA, Troshin PA, Peregudov AS, Kozlovskiy VI, Balzarini J, Lyubovskaya RN (2007) *Org Biomol Chem* 5:2783
- Promsri S, Chuichaya P, Sanghiranb V, Parasuka V, Hannongbua S (2005) *J Mol Struct THEOCHEM* 715:47
- Durdagi S, Mavromoustakos T, Chronakis N, Papadopoulos MG (2008) *Bioorg Med Chem* 16:9957
- Durdagi S, Mavromoustakos T, Papadopoulos MG (2008) *Bioorg Med Chem Lett* 18:6283
- Durdagi S, Supuran CT, Strom A, Doostdar N, Kumar MK, Barron AR, Mavromoustakos T, Papadopoulos MG (2009) *J Chem Inf Model* 49:1139
- Waterbeemd H, Gifford E (2003) *Nat Rev Drug Discov* 2:192
- ArgusLab 4.0.1 Mark A. Thompson Planaria Software LLC, Seattle, WA [www.arguslab.com](http://www.arguslab.com)
- Rappe AK, Casewit CJ, Colwell KS, Goddard WA III, Skiff WM (1992) *J Am Chem Soc* 114:10024
- Broyden CG (1970) 6:76–90; Fletcher R (1970) 13:317–322; Goldfarb DA (1970) 24:23–26; Shanno DF (1970) *Math Comput* 24:647
- Leach AR (2001) *Molecular modelling: principles and applications*. 2nd Edn. Pearson Education Ltd
- Sengupta D, Verma D, Naik PK (2007) *J Biosci* 32:1316
- SYBYL 8.0, Tripos International, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA
- Case DA, Cheatham T, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods R (2005) *J Comput Chem* 26:1668
- Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Merz M, Wang B, Hayik S, Roitberg SA, Seabra I, Wong F, Paesani J, Vanicek X, Wu SR, Brozell V, Tsui H, Gohlke L, Yang C, Tan J, Mongan V, Hornak G, Cui P, Beroza DH, Mathews C, Schafmeister WS, Ross, Kollman PA (2006) AMBER 9, University of California, San Francisco
- RSCB Protein Databank; <http://www.rcsb.org/pdb/home/home.do>
- Hyland LJ, Tomaszek TA Jr, Roberts GD, Carr SA, Magaard VW, Bryan HL, Fakhoury SA, Moore ML, Minnich MD (1991) *Biochemistry* 30:8454
- Zhu Z, Schuster DI, Tuckerman ME (2003) *Biochemistry* 42:1326
- Pietrucci F, Marinelli F, Carloni P, Laio A (2009) *J Am Chem Soc* 131:11811
- Piana S, Sebastiani D, Carloni P, Parrinello M (2001) *J Am Chem Soc* 123:8730
- Hou T, McLaughlin WA, Wang W (2008) *Proteins* 71:1163
- Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) *Proteins* 65:712
- Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) *J Comput Chem* 25:1157
- Jakalian A, Bush BL, Jack DB, Bayly CI (2000) *J Comput Chem* 21:132
- Jakalian A, Jack DB, Bayly CI (2002) *J Comput Chem* 23:1623
- Jorgensen WLCJ, Madura JD, Impey RW, Klein ML (1983) *J Chem Phys* 79:926
- Darden T, York D, Pedersen L (1993) *J Chem Phys* 98:10089
- Ryckaert JP, Ciccotti G, Berendsen HJC (1977) *J Comput Phys* 23:327
- Izaguirre JA, Catarello DP, Wozniak JM, Skeel RD (2001) *J Chem Phys* 114:2090
- Kollman PA et al (2000) *Acc Chem Rev* 33:889
- Sitkoff D, Sharp KA, Honig B (1998) *J Phys Chem* 98:1978
- Srinivasan J, Cheatham TE, Cieplak P, Kollman PA (1998) *J Am Chem Soc* 120:9401
- Wang W, Donini O, Reyes CM, Kollman PA (2001) *Annu Rev Biophys Biomol Struct* 30:211
- Gouda H, Kuntz ID, Case DA, Kollman PA (2003) *Biopolymers* 68:16
- Stoica I, Sadiq SK, Coveney PV (2008) *J Am Chem Soc* 130:2639
- Hou T, Wang J, Li Y, Wang W (2011) *J Chem Inf Model* 51:69
- QikProp, version 3.3, Schrödinger, LLC, New York, NY, 2009
- Yazdaniyan M, Glynn SL, Wright JL, Hawi A (1998) *Pharm Res* 15:1490

52. Irvine JD, Takahashi L, Lockhart K, Cheong J, Tolan JW, Selick HE, Grove JR (1999) *J Pharm Sci* 88:28
53. Durdagi S, Subbotina J, Lees-Miller J, Guo, J, Duff HJ, Noskov SY (2010) *Curr Med Chem* 17:3514
54. Ajay, Bemis GW, Murcko MA (1999) *J Med Chem* 42:4942
55. Friedman HS, Ganapathi PS, Rubin PS, Kenyon GL (1998) *J Med Chem* 41:2424
56. Bingel C (1993) *Chem Ber* 126:1957
57. Ganapathi PS, Friedman SH, Kenyon GL, Rubin YJ (1995) *Org Chem* 60:2954
58. Freedberg DJ, Ishima R, Jacob J, Wang YX, Kustanovich I, Louis JM, Trochia DA (2002) *Protein Sci* 11:221
59. Zoete V, Michielin O, Karplus M (2002) *J Mol Biol* 315:21
60. Reyes CM, Kollman PA (2000) *J Mol Biol* 297:1145
61. Wang CW, Lim WA, Jakalian A, Wang J, Wang M, Luo R, Bayly CT, Kollman PA (2001) *J Am Chem Soc* 123:3986
62. Hou T, Yu R (2008) *J Med Chem* 50:1177
63. Cai Y, Schiffer CA (2010) *J Chem Theory Comp* 6:1358
64. Charlier L, Nespoulous C, Fiorucci S, Antonczak S, Golebiowski G (2007) *Phys Chem Chem Phys* 9:576
65. <http://www.gene.com/gene/products/information/invirase/>
66. Ioakimidis L, Thoukydidis L, Mirza A, Naeem S, Reynisson J (2008) *QSAR Comb Sci* 4:445
67. Luo S, Wang Z, Patel M, Khurana V, Zhu X, Pal D, Mitra AK (2011) *Int J Pharm* 414:77
68. Anson BD, Weaver JGR, Ackerman MJ, Akinsete O, Henry K, January CT, Badley AD (2005) *Lancet* 365:682
69. Ohtaka H, Velazquez-Campoy A, Xie D, Freire E (2002) *Protein Sci* 11:1908
70. Chen Z, Li Y, Cheng E, Hall DL, Darke PL, Culbertson C, Shafer JA, Kuo LC (1994) *J Biol Chem* 269:26344
71. Kaldor SW, Kailsah VJ, Davies JF II, Shetty BV, Fritz JE, Appelt K, Burgess JA, Campanale KM, Chirgadze NY, Clawson DK, Dressman BA, Hatch SD, Khalil DA, Kosa MB, Lubbehusen PP, Muesing MA, Patick AK, Reich SH, Su KS, Tatlock JH (1997) *J Med Chem* 40:3797
72. Mongan J, Case DA, McCammon JA (2004) *J Comput Chem* 25:2038
73. Lefebvre E, Schiffer CA (2008) *AIDS Rev* 10:131–142
74. Muzammil S, Armstrong AA, Kang LW, Jakalian A, Bonneau PR, Schmelmer V, Amzel LM, Freire E (2007) *J Virol* 81:5144
75. Ioannou E, Hirsch A, Elemen Y (2007) *Tetrahedron* 63:7070
76. Giordani S, Colomer JF, Cattaruzza F, Alfonsi J, Meneghetti M, Prato M, Bonifazi D, Carbon, (2009) 47:578
77. Tasis D, Tagmatarchis N, Georgakilas V, Gamboz C, Soranzo MR, Prato M (2003) *C. R. Chimie* (2003) 6:597
78. Bjelakovic MS, Godjevac DM, Milic DR (2007) *Carbon* 45:2260
79. Tsumoto H, Takahashi K, Suzuki T, Nakagawa H, Kohda K, Miyata N (2008) *Bioorg Med Chem Lett* 18:657
80. Huang ST, Ho CS, Lin CM, Fang HW, Peng YX (2008) *Bioorg Med Chem* 16:8619
81. de Freitas RP, Iehl J, Delavaux-Nicot B, Nierengarten JF (2008) *Tetrahedron* 64:11409
82. Thompson BC, Frechet JM (2008) *Angew Chem Int Ed* 47:58
83. Loboda O, Zalesny R, Avramopoulos A, Luis JM, Kirtman B, Tagmatarchis N, Reis H, Papadopoulos MG (2009) *J Phys Chem A* 113(6):1159
84. Zalesny R, Loboda O, Iliopoulos K, Chatzikyriakos G, Couris S, Rotas G, Tagmatarchis N, Avramopoulos A, Papadopoulos MG (2010) *Phys Chem Chem Phys* 12:373