Bioorganic & Medicinal Chemistry Letters 18 (2008) 6283–6289

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

# 3D QSAR CoMFA/CoMSIA, molecular docking and molecular dynamics studies of fullerene-based HIV-1 PR inhibitors

Serdar Durdagi [a,b,*], Thomas Mavromoustakos [a,c], Manthos G. Papadopoulos [a,*]

[a] Institute of Organic and Pharmaceutical Chemistry, The National Hellenic Research Foundation, Vas. Constantinou Ave. 48, 11635 Athens, Greece
[b] Department of Biology Chemistry and Pharmacy, Freie Universität Berlin, Takustr. 3, 14195 Berlin, Germany
[c] Department of Chemistry, University of Athens, 15784 Athens, Greece

ARTICLE INFO

*Article history:*
Received 4 August 2008
Revised 11 September 2008
Accepted 11 September 2008
Available online 8 October 2008

*Keywords:*
Fullerene
3D QSAR
CoMFA
CoMSIA
Molecular docking
Molecular dynamics
HIV-1 PR

ABSTRACT

For the first time, a set of experimentally reported [60] fullerene derivatives were subjected to the 3D-QSAR/CoMFA and CoMSIA studies. The aim of this study is to propose a series of novel [60] fullerene-based inhibitors with optimal binding affinity for the HIV-1 PR enzyme. The position of the template molecule at the cavity of HIV-1 PR was optimized and 3D QSAR models were developed. Relative contributions of steric/electrostatic fields of the 3D-QSAR/CoMFA and CoMSIA models have shown that steric effects govern the bioactivity of the compounds, but electrostatic interactions play also an important role.The de novo drug design Leapfrog simulations provided a series of novel compounds with predicted improved inhibition effect.



Inhibition of human immunodeficiency virus type 1 aspartic protease (HIV-1 PR) leads to the production of immature virus particles and prevention of further rounds of infection.[1,2] HIV-1 PR is an important target enzyme for anti-acquired immunodeficiency syndrome (AIDS) drug design as its inhibition leads to the production of non-infectious viral particles.[1] In the last few years, fullerene and its derivatives have been extensively investigated for biomedical applications. Inhibition of HIV-1 PR by fullerene analogues was demonstrated by Friedman et al.[3] and complexation of HIV-1 PR with fullerene compounds has been supported by molecular modeling studies.[4] These studies showed that the fullerene can be perfectly accommodated inside the binding pocket of HIV-1 PR. The binding affinity values of 'first generation' fullerene inhibitors were not significant ($EC_{50} \sim 10^{-6}$ M). Thus, further structural investigation is required in order to propose new HIV-1 PR/fullerene complexes with better binding affinity. For this aim, a set of biologically evaluated synthetic fullerene derivatives (Table 1)[4–8] have been used to construct models performing three-dimensional quantitative structure–activity relationships (3D-QSAR) methodologies: comparative molecular field analysis (CoMFA)[9] and comparative molecular similarity indices analysis (CoMSIA).[10] The selection of the putative bioactive lowest energy conformation of the template molecule and the superimposition of all molecules on template compound are the two most critical steps in the 3D QSAR CoMFA and CoMSIA studies.[11] The availability of a 3D model of a receptor improves the structure alignment and can provide statistically more reliable models.[11] Therefore, the most potent fullerene analogue up to date (compound **4** in training set, Table 1) has been selected as template molecule and it is docked in the binding cavity of 3D model of HIV-1 PR. The derived best docked complex structure has been subjected to molecular dynamics (MD) simulations, in order to stabilize the localization of fullerene at the HIV-1 PR. 3D-QSAR/CoMFA and CoMSIA methods were applied to the data set, which was divided into training and test sets. To our knowledge, this study is the first 3D-QSAR application to the fullerene-based compounds. Both CoMFA and CoMSIA studies gave similar results indicating that the steric effects are essential for the activity. The de novo design Leapfrog routine of SYBYL[12] was used as an aid for the discovery of new molecules based on the 3D QSAR results.

More details follow the afore-mentioned procedures used and the significance of the results obtained upon using them. The FlexX docking program of SYBYL molecular modeling package[12] has been used for docking simulations. The default FlexX scoring function was used through the simulations. FlexX uses a fast docking method that allows flexibility in the ligands, keeping the receptor rigid. FlexX uses formal charges, which were turned on during the docking. The scoring function of FlexX which was developed by Böhm[13]

* Corresponding authors. Tel.: +30 2107273894; fax: +30 2107273872.
*E-mail addresses:* durdagis@eie.gr (S. Durdagi), mpapad@eie.gr (M.G. Papadopoulos).

0960-894X/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2008.09.107

**Table 1**
Chemical structures and experimental (observed) binding affinities of fullerene derivatives at HIV-1 PR for 3D QSAR/CoMFA and CoMSIA studies[4–8]

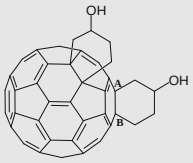
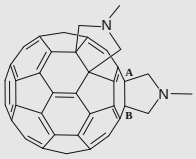
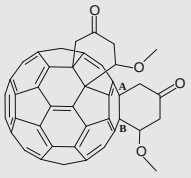
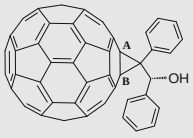
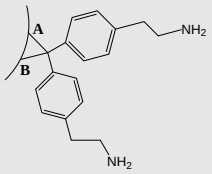
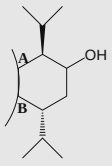
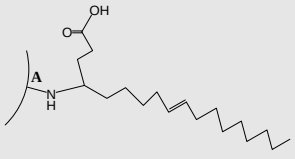
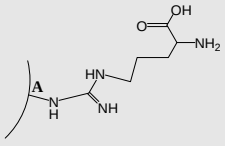
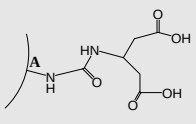
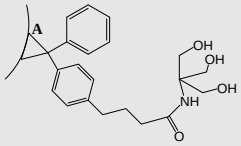
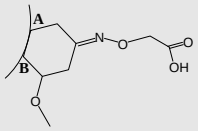
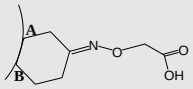
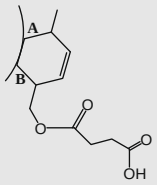
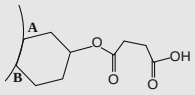
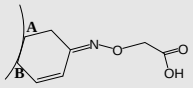
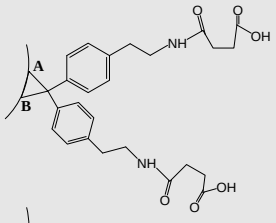
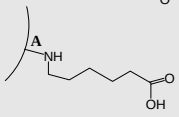
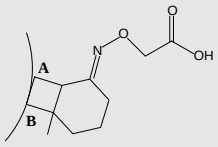
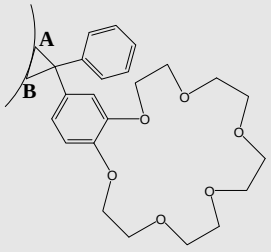
| Compound | Structure | $EC_{50}$ for HIV-1 (μM) | Obs. $pEC_{50}$ | $pEC_{50}$ (CoMFA) | $pEC_{50}$ (CoMSIA) |
|---|---|---|---|---|---|
| **1**[a] |  | 17.6 | 4.75 | 4.64 | 4.73 |
| **2**[a] |  | 72.7 | 4.14 | 4.17 | 4.15 |
| **3**[a] |  | 7.70 | 5.11 | 5.17 | 5.18 |
| **4** |  | 0.10 | 7.0 | 7.05 | 6.95 |
| **5** |  | 5.0 | 5.3 | 5.25 | 5.31 |
| **6** |  | 0.15 | 6.8 | 6.65 | 6.84 |
| **7** |  | 1300 | 2.89 | 2.84 | 2.97 |
| **8** |  | 75.0 | 4.12 | 4.03 | 4.08 |
| **9** |  | 230 | 3.64 | 3.69 | 3.56 |
| **10** |  | 140.0 | 3.85 | 3.92 | 3.80 |

**Table 1** (*continued*)

| Compound | Structure | $EC_{50}$ for HIV-1 (μM) | Obs. $pEC_{50}$ | $pEC_{50}$ (CoMFA) | $pEC_{50}$ (CoMSIA) |
|---|---|---|---|---|---|
| **11** |  | 2.50 | 5.60 | 5.52 | 5.62 |
| **12** |  | 0.9 | 6.05 | 5.96 | 5.94 |
| **13** |  | 2.9 | 5.53 | 5.59 | 5.61 |
| **14** |  | 2.2 | 5.66 | 5.83 | 5.66 |
| **15** |  | 6.30 | 5.20 | 5.14 | 5.16 |
| **16** |  | 7.30 | 5.14 | 5.23 | 5.23 |
| **17** |  | 7.3 | 5.13 | 5.22 | 5.11 |
| **18** |  | 0.49 | 6.31 | 5.22 | 5.76 |
| **19** |  | 21.7 | 4.66 | 4.68 | 4.67 |
| **20** |  | 137 | 3.86 | 3.46 | 3.67 |

Only pendant groups have been shown in the table for mono adduct fullerenes, except **4** (template compound). Predicted $pEC_{50}$ values obtained by CoMFA and CoMSIA have also been shown in the last two columns.

[a] Schuster et al.[7] noted that the regiochemistry of **1**-**3** is undetermined. For the present computations, we assumed these structures which have been given in this table.

in order to rank the docking mode solutions, provides an estimate of the free binding energy $\Delta G$ of the protein–ligand complexes. The best scored docked mode of template molecule **4** at HIV-1 PR was used as input for MD simulations. Using MD simulations, position of **4** at the catalytic site of HIV-1 PR was optimized. This technique assists the improvement of alignment procedure in 3D QSAR.

The MD simulations were performed with GROMACS 3.3.1 software package[14] employing the gmx force field.[15] Canonical NVT ensemble at 300 K was used with periodic boundary conditions, and the temperature was kept constant by the Berendsen thermostat.[16] Electrostatic interactions were calculated using the particle mesh Ewald method.[17] Cut-off distances for the calculation of Coulomb and van der Waals interactions were 1.0 and 1.4 nm, respectively. Prior to the dynamics simulation, energy minimization was used to the full system ([HIV-1 PR/template molecule] complex and solvent (water) molecules) (Fig. 1) without constraints using the steepest descent integrator for 5000 steps. The system was then equilibrated via a 100 ps MD simulations at 300 K. Finally, a 2-ns simulation was performed with a time step of 2 fs.

The flexibility of the flap region of the HIV-1 PR enzyme has been discussed using both experimental and computational studies.[18,19] The free protease adopts closed, semi-open or fully open forms in a dynamic equilibrium. However, the closed form is favored when the inhibitor is bound in the cavity of the reaction site.[18] It has been proposed by Friedman et al.[3,4] that there is a direct correlation between the binding affinity of an inhibitor and the amount of hydrophobic surface area that it can desolvate. The MD simulations by Zhu et al.[20] showed the exclusion of water near the flap regions in order to accommodate the fullerene inhibitor. The reduction of the water density in the cavity leads to the enhancement of the hydrophobic interaction between the fullerene derivative and the active site of the receptor. Trajectory analysis results showed that the flaps of HIV-1 PR were pulled toward the bottom of active site and obtained a closed form. The decreasing area of the binding pocket during the simulation has been represented by changing the distance between $C^{\alpha}$ atoms of the four critical amino acids: ASP25-ASP25′, GLY49-GLY49′, ASP25-ASP49, and ASP25′-GLY49′ (Fig. 2, top). All of the calculated distances have been decreased gradually except the distance between $C^{\alpha}$ atoms of ASP25-GLY49, which is kept constant throughout the simulation (Fig. 2, bottom). Therefore, the MD simulations have shown that inhibitor leads to a closed form of HIV-1 PR throughout the simulation.

After obtaining the optimum position of **4** inside the HIV-1 PR, the localization of the rest of the data set in Table 1 has been defined based on the template compound and 3D QSAR/CoMFA and CoMSIA methods have been developed. The aim of applying the

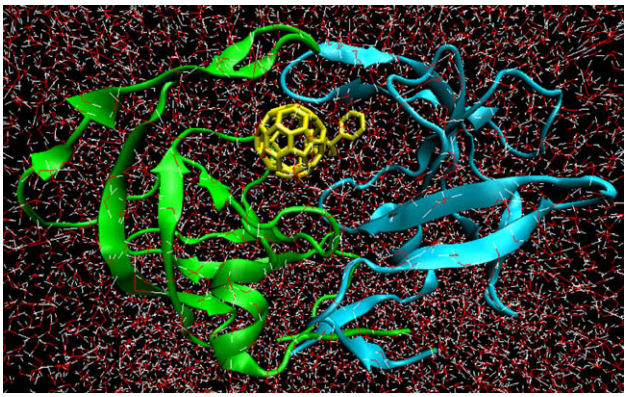


**Figure 1.** Full system [HIV-1 PR/template molecule] complex and solvent (water) molecules used in MD simulations.

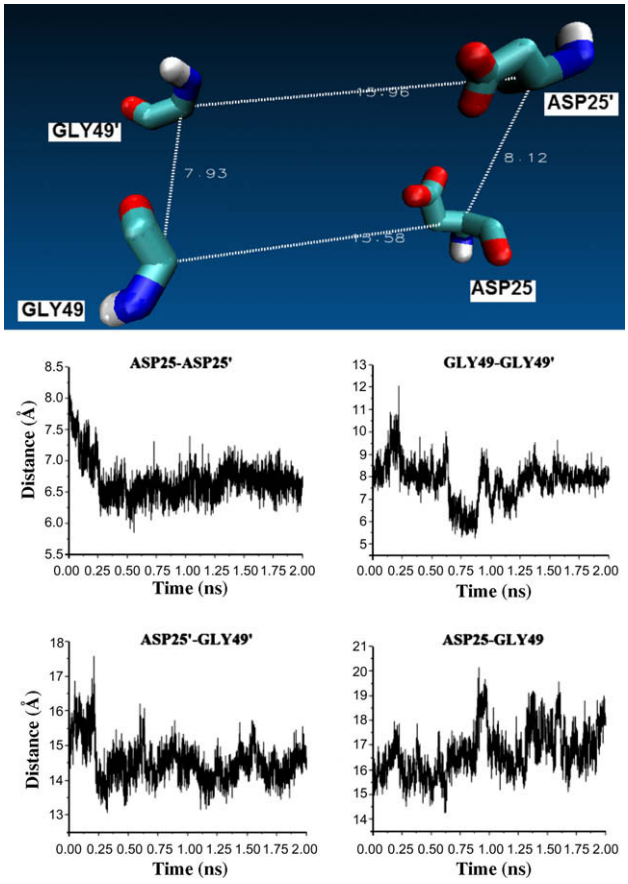



**Figure 2.** Top: Decreasing of binding cavity during the simulation has been represented by changing the distance between $C^{\alpha}$ atoms of the four critical amino acids. These are ASP25, ASP25′, GLY49 and GLY49′. Bottom: Change of distances throughout simulation.

3D QSAR/CoMFA and CoMSIA methodologies is to derive indirect binding information from the correlation between the bioactivity of a training set of compounds and their 3D structures. We have previously showed that docking experiments along with 3D QSAR studies constitute a powerful tool in the rational drug design since they lead to more reliable statistical models.[11,21] The logarithmic 1/$EC_{50}$ values ($pEC_{50}$) were used in the 3D-QSAR correlations, as they are related to changes in the free energy of binding. Several variations in the alignment procedures are considered by superimposing similar pharmacophoric features. Highlighted carbon atoms (32 central carbon atoms of fullerene) at the template ligand **4** are selected for the structural superimposition processes (Supporting Information Fig. 1i). The alignment of the molecules was based on atom-by-atom superimposition of selected atoms, which are common in all compounds. The superimposition of the molecules in the training set has been shown in Supporting Information Figure 1ii.

Compounds **1–17** have been used as training set and **18–20** used as a test set. Test set compounds are selected from the data set to possess higher, similar, and lower inhibition effects than average $pEC_{50}$ values of the studied compounds which represent the whole data set. The cross validated partial-least square (PLS) method was then subjected to the training set. Table 2 summarizes the statistical results.

3D-QSAR/CoMFA and CoMSIA studies gave cross validated $r^2$ ($r^2_{cv}$) values of 0.549 and 0.555, respectively. The non-cross validated PLS analysis yielded $r^2$ values of 0.994 and 0.997, respectively. These results satisfy the statistical validity criteria and

**Table 2**
Statistical results of 3D QSAR/CoMFA and CoMSIA models

| | CoMFA | CoMSIA |
|---|---|---|
| $r^2_{cv}$ | 0.549 | 0.555 |
| $r^2$ | 0.994 | 0.997 |
| Standard error of estimate | 0.097 | 0.051 |
| $F$ | 509.545 | 1216.442 |
| Relative contributions of steric/electrostatic fields | 0.776:0.224 | 0.872:0.128 |
| Number of optimal components | 4 | 5 |

allow for the construction of a significant QSAR model. Predicted $PEC_{50}$ values of compounds by constructed CoMFA and CoMSIA methods have been given in Table 1. Figure 2 in Supporting Information shows the experimental versus predicted $PEC_{50}$ values for CoMFA and CoMSIA. The significance of the proposed models is verified by the good predictions of the activity of compounds belonging to the test set. Test set compounds **19** and **20** both in CoMFA and CoMSIA gave good predictions (Table 1). $pEC_{50}$ value of compound **18** is underestimated by CoMFA model about 1.0 unit but CoMSIA result gave satisfactory estimation of its activity (error on prediction is less than a unit).

The CoMSIA and its precursor CoMFA are leading to high quality graphical maps which describe mainly the steric and the electrostatic requirements of the binding cavity of HIV-1 PR in the 3D Cartesian space. Figure 3 shows the steric-electrostatic contour maps of the CoMFA and CoMSIA models for the compounds that show the best and worst inhibition effects within the data set for the HIV-1 PR receptor (compounds **4** and **7** in Table 1, respectively).

The individual contributions of the steric and the electrostatic favored and disfavored levels are initially fixed at 80% and 20%, respectively. In order to show clearly the stereoelectronic contours, in a subsequent simulation, steric and electrostatic favored and disfavored levels are modified at 90% and 10%, respectively, for CoMFA plots (the corresponding figure has been presented in Fig. 3 of Supporting Information). The contours for steric fields are shown in green (bulky groups favored) and yellow (bulky groups not favored) colors, while the electrostatic field contours are shown in red (electronegative substituents favored) and blue (electropositive substituents favored) colors. Contour plots confirmed the stability of the constructed models. For example, **4** has better inhibition effect than **7**, around $10^4$-fold. This difference can be explained by different topographical requirements of **4** and **7**; both aromatic rings of **4**, fit very well within the green colored contour map which shows sterically favorable places, there are no bulky groups of **7** in this field (Fig. 3). In addition, inactive compound **7** has a flexible chain and fits within the yellow colored contour which shows sterically unfavorable areas at the CoMSIA model (Fig. 3, right). Moreover, electrostatic contour maps of **4** and **7** clearly show and explain the differences of bioactivities of **4** and **7** both in CoMFA and CoMSIA models. For example, the –OH group of **4** fits within the red colored contour which shows negative potential favored fields while, –C=O group of **7** fits within the blue colored contour which shows negative potential unfavored (Fig. 3) regions. Relative contributions of steric and electrostatic fields obtained by CoMFA and CoMSIA are 0.78:0.22 and 0.87:0.13, respectively. Therefore, models have shown that steric effects are important for the bioactivity of the compounds; however, electrostatic interactions have also a contribution which should be taking into account in the biological action.

The second generation de novo drug discovery module Leapfrog of SYBYL molecular modeling package[12] was used in order to aid the design of a series of new potentially active fullerene-based HIV-1 PR inhibitors. Leapfrog performs by repeating various structural changes on a pre-chosen lead molecule and then keeps only the results satisfying the steric and chemical restraints of the binding site and discards the unsuccessful analogues depending on the binding energy results.[22,23] As an initial basic procedure of Leapfrog a virtual receptor site was generated using the obtained CoMFA model. The template molecule **4** was used as starting structure. The mode OPTIMIZE, suggesting improvements to the existing lead, was used in the present work, with the additional constraint of minimizing synthetic difficulties during the seeking procedure. JOIN, FUSE, WEED, CROSSOVER modules were employed after the initial run of 100 moves and the derived ligands that had the best binding energy were used for the repeating cycle of 5000 moves. The proposed molecules produced by Leapfrog and their predicted activities are shown in Table 3. Most of the proposed compounds have a significant improvement in the binding affinity in comparison to the template compound **4** which has the best currently available experimental value for fullerene-based HIV-1 PR inhibitors.

In conclusion, experimentally reported [60] fullerene analogues were used to construct 3D QSAR models in order to optimize the

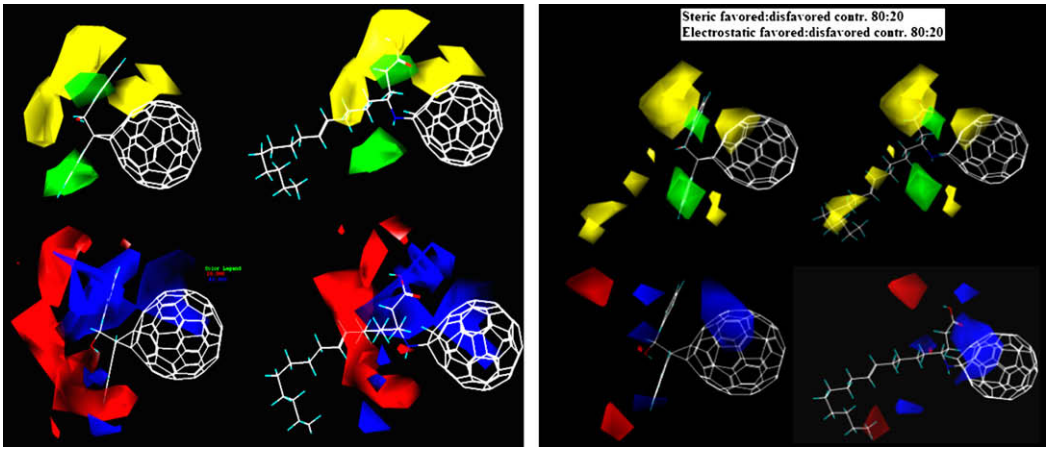



**Figure 3.** Left: CoMFA contour maps of template compound **4** (shown in top-left and bottom-left which has highest binding affinity in the training set) and compound **7** (shown in top-right and bottom-right which has lowest binding affinity in the training set). Sterically favored areas are shown in green color (contribution level of 80%). Sterically unfavored areas are shown in yellow color (contribution level of 20%). Positive potential favored areas are shown in blue color (contribution level of 80%). Positive potential unfavored areas are shown in red color (contribution level of 20%). Right: CoMSIA contour maps of template compound **4** (shown in top-left and bottom-left) and compound **7** (shown in top-right and bottom-right). Same steric/electrostatic contribution of CoMFA has been kept for the CoMSIA.

**Table 3**
The proposed molecules produced by Leapfrog and their predicted $pEC_{50}$ values using CoMFA model. For clarity, the reference compound (template) **4** has been shown with pale color and edited part has been shown with dark color

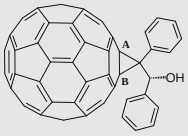
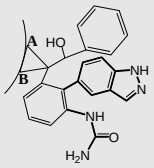
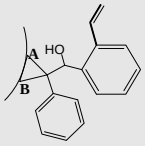
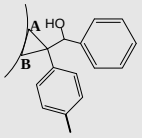
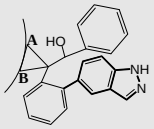
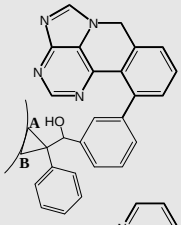
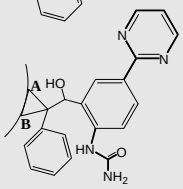
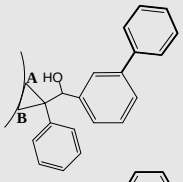
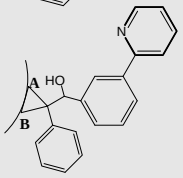
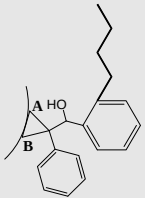
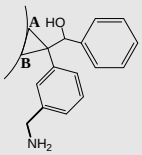
| Compound | Structure | $pEC_{50}$ |
|---|---|---|
| Ref. |  | 7.05[a] |
| LPF_1 |  | 7.87 |
| LPF_2 |  | 7.87 |
| LPF_3 |  | 7.80 |
| LPF_4 |  | 7.73 |
| LPF_5 |  | 7.72 |
| LPF_6 |  | 7.58 |
| LPF_7 |  | 7.50 |
| LPF_8 |  | 7.36 |
| LPF_9 |  | 7.26 |

**Table 3** (*continued*)

| Compound | Structure | $pEC_{50}$ |
|---|---|---|
| LPF_10 |  | 7.14 |

[a] Predicted $pEC_{50}$ by CoMFA model.

binding energy of HIV-1 PR/fullerene complexes, and to propose novel bioactive compounds. In order to improve the superimposition procedure of 3D-QSAR/CoMFA and CoMSIA methodologies, molecular docking and MD simulations have been used. PLS results and contour maps confirm the stability of constructed models. Relative contributions of steric/electrostatic fields have shown that steric effects are essential for the bioaffinity, but electrostatic interactions play also an important role in the biological action. Leapfrog simulations provided novel compounds with improved inhibition effect. The molecular characteristics of structures generated by Leapfrog fit very well with generated stereoelectronic contour maps. To our knowledge, the present study is the first application to date for 3D-QSAR of fullerene-based molecules. The obtained models can serve as a basis for the design of novel fullerene-based HIV-1 PR inhibitors with enhanced activity.

## Acknowledgments

This work was financially supported by the European Union within 6th Framework Programme-Marie Curie Actions (Project: EURODESY-MEST-CT-2005-020575).

## Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2008.09.107. Selected atoms of the template compound **4** for structural superimpositions of the compounds in training set. Structural alignments of the compounds in the training set. Plots of observed and 3DQSAR/CoMFA and CoMSIA-predicted binding affinities (given as pEC50) of fullerene analogues in the training set at the HIV-1 PR and CoMFA contour maps of compounds **4** and **7** using contributions level 90% and 10% for favorable and unfavorable areas, respectively.

## References and notes

1. Toth, G.; Borics, A. *J. Mol. Graph. Model.* **2006**, *24*, 465.
2. Bosi, S.; Da Ros, T.; Spalluto, G.; Balzarini, J.; Prato, M. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 4437.
3. Friedman, H. S.; DeCamp, D. L.; Sijbesma, R. P.; Srdanov, G.; Wudl, F.; Kenyon, G. *J. Am. Chem. Soc.* **1993**, *115*, 6506.
4. Friedman, H. S.; Ganapathi, P. S.; Rubin, P. S.; Kenyon, G. L. *J. Med. Chem.* **1998**, *41*, 2424.
5. Bingel, C. *Chem. Ber.* **1993**, *126*, 1957.
6. Ganapathi, P. S.; Friedman, S. H.; Kenyon, G. L.; Rubin, Y. *J. Org. Chem.* **1995**, *60*, 2954.
7. Schuster, D. I.; Wilson, S. R.; Schinazi, R. F. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 1253.
8. http://chemdb.niaid.nih.gov.
9. Cramer, R. D. I.; Depriest, S.; Patterson, D.; Hecht, P. The Developing Practice of Comparative Molecular Field Analysis; Kubinyi, H., Ed.; ESCOM: Leiden, 1993, p 443.
10. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130.
11. Durdagi, S.; Papadopoulos, M. G.; Papahatjis, D. P.; Mavromoustakos, T. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 6754.
12. Sybyl Molecular Modeling Software Packages, ver. 6.8., 2001, Tripos Inc., St Louis, MO.
13. Böhm, H. J. *J. Comput-Aided Mol. Des.* **1994**, *10*, 427.
14. van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J. Comput. Chem.* **2005**, *26*, 1701.

15. Lindhal, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306.
16. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Comput. Chem.* **1984**, *81*, 3684.
17. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
18. Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *Proc. Nat. Am. Sci.* **2006**, *103*, 915.
19. Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *J. Am. Chem. Soc.* **2006**, *128*, 2812.
20. Zhu, Z.; Schuster, D. I.; Tuckerman, M. E. *Biochemistry* **2003**, *42*, 1326.
21. Durdagi, S.; Reis, H.; Papadopoulos, M. G.; Mavromoustakos, T. *Bioorg. Med. Chem.* **2008**, *16*, 7377–7387.
22. Nair, C. P.; Sobiha, M. E. *J. Mol. Graph. Model.* **2007**, *26*, 117.
23. Makhija, M. T.; Kasliwal, R. T.; Kulkarni, V. M.; Neamati, N. *Bioorg. Med. Chem.* **2004**, *12*, 2317.