Original article

Docking-based 3D-QSAR study of HIV-1 integrase inhibitors

Pawan Gupta [a], Nilanjan Roy [a,b], Prabha Garg [a,*]

[a] Centre for Pharmacoinformatics, National Institute of Pharmaceutical Education and Research (NIPER), Sector-67, S.A.S. Nagar, 160062 Punjab, India
[b] Department of Biotechnology, National Institute of Pharmaceutical Education and Research (NIPER), Sector-67, S.A.S. Nagar, 160062 Punjab, India

ARTICLE INFO

ABSTRACT

In this study, 3-aroyl-1,1-dioxo-1,4,2-benzodithiazine and 4-chloro-N-(4-oxopyrimidin-2-yl)-2-mercaptobenzenesulfonamide derivatives (HIV-1 integrase inhibitors) were used for CoMFA and CoMSIA to determine the substructures required for the activity of these molecules. To explore the binding mode of inhibitors, docking studies were done and docked conformation of highly active molecule was used as template for alignment. The best CoMFA model yielded the cross validation $r^2_{cv} = 0.728$, non-cross validation $r^2_{ncv} = 0.934$ and predictive $r^2_{pred} = 0.708$. The best CoMSIA model yielded a cross validation $r^2_{cv} = 0.794$, non-cross validation $r^2_{ncv} = 0.928$ and predictive $r^2_{pred} = 0.59$. It was found that steric (CoMFA) and hydrophobic fields (CoMSIA) have large contribution towards the inhibitory activity than the other fields. Docking and 3D-QSAR studies have provided clues to a better understanding of interaction between the inhibitors and HIV-1 integrase.

© 2009 Elsevier Masson SAS. All rights reserved.

1. Introduction

Human immunodeficiency virus (HIV) causes acquired immunodeficiency syndrome (AIDS) [1–5]. Three enzymes are essential for replication of this virus inside host, reverse transcriptase (RT), protease (PR) and integrase (IN). All of them are considered to be promising targets for the development of anti-HIV drugs [1,2,4,6]. The intense researches on RT and PR have been done for development of new therapeutic agents against AIDS [5]. Combination of RT and PR inhibitors developed for therapeutic intervention has limited efficacy because of side effect and drug resistance [5,7]. Thus, current situation warrants search for new anti-HIV agents with fewer side effects and high efficacy. HIV-1 IN has been recognized as another lucrative target for inhibition of viral replication. HIV-1 IN catalyzes the integration of viral genome into host chromosome through a two-step mechanism. First step is 3′-processing while the second step is strand transfer. In the first step, IN removes two nucleotides from each 3′ end of the viral DNA. The 3′ processed strands are then incorporated to host DNA [5–8]. Although RT and PR orthologues are present in humans, no orthologue of HIV-1 IN is reported till date, thereby allowing selective drug design [9]. This is one of the reasons for choosing the IN as an attractive and validated target for AIDS.

Mn^{2+} or Mg^{2+} cofactor is required for IN activity [10]. IN activity is optimum in presence of Mn^{2+}, however Mg^{2+} is abundantly present in living cells, thus Mg^{2+} is physiologically relevant cation for HIV-1 IN activity [11]. It has been hypothesised that ortho-hydroxy acid group of styrylquinoline derivatives are possibly involved in coordination of metal ion [12]. This hypothesis was reinforced by computational studies of styrylquinoline derivatives with RSV IN, energy minimized top binding interactions were found near the Mg^{2+} in the vicinity of IN active site [13]. Moreover, it has also been reported that interaction between acid moiety of diketo acid and metal ion causes chelation of metal cofactor activity in active site of IN. As a result of this chelation, the activity of IN is abolished [14].

IN comprises of three domains, (1) N-terminal domain (residues 1–50), contains His_2Cys_2 motif, it coordinates Zn^{2+}; required for promoting multimerization and increasing the catalytic activity [15,16], (2) Highly conserved catalytic core domain (residues 51–212), contains catalytic triad motif [D,D (35) E] of three essential acidic residues responsible for disintegration reaction. Metal is coordinated by these acidic residues and plays an important role in catalysis. Mutation at core domain abolishes the catalytic activity [17,18]. This domain defines the position of active site [19]. (3) C-terminal domain (residues 213–288) binds DNA nonspecifically. 3′-processing and strand transfer reaction is performed by both C-terminal and N-terminal domains, so both the domains constitute important structural part for integrase activity [10,17,20].

Generally, for Quantitative Structure Activity Relationship (QSAR) model building only those molecules are selected which

* Corresponding author. Tel.: +91 172 2214 682; fax: +91 172 2214 692.
E-mail addresses: prabhagarg@niper.ac.in, gargprabha@yahoo.com (P. Garg).

have exact biological activity value. However, a good model can also be developed by using those molecules for which, only some range (e.g. $IC_{50} > 100 \, \mu M$) is given instead of exact value of activity. Definitely some structural information is associated with these molecules which make them inactive. So this information should also be taken into account for understanding the QSAR model of those molecules. This can ensure that the QSAR model can predict activity accurately also for those molecules which are within the range of training set molecules [21]. Here, we considered molecules of $>100 \, \mu M$ IC_{50} value equal to $100 \, \mu M$ IC_{50}.

Over the last few years, QSAR and docking studies have been carried out for different series of HIV-1 IN inhibitors. Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA), three-dimensional quantitative structure activity relationship (3D-QSAR) techniques, had been performed on HIV-1 IN inhibitors, belonging to 11 structurally different classes [22]. On the other hand, Kuo et al. performed 3D-QSAR and docking studies with objective of optimization of mercaptobenzenesulfonamide as HIV-1 IN inhibitors [23].

Here, we describe the QSAR studies of 3-aroyl-1,1-dioxo-1,4,2-benzodithiazines and 4-chloro-N-(4-oxopyrimidin-2-yl)-2-mercaptobenzenesulfonamide derivatives which act as HIV-1 IN inhibitors [24,25]. To explore the binding mode of inhibitors, docking studies were done on crystal structure of HIV-1 IN (PDB ID 1QS4) and docked conformation of highly active molecule was used as template for alignment. PDB ID 1QS4 is widely used for computational studies, because only 1QS4 has co-crystallized ligand in their crystal structure but others (PDB ID 1BL3, BIS) do not have any co-crystallized ligand [9,26]. As evident from the reported literature [27–31], combined use of docking study with 3D-QSAR can provide more information on the interaction mode between agonist and the receptor.

There are two objectives of these studies. First, to explore the binding mode of this agonist with HIV-1 IN enzyme. Second, to develop 3D-QSAR models for aforementioned structural classes of HIV-1 IN inhibitors and to examine the structure activity relationship for the HIV-1 IN inhibitory activity of these inhibitors using CoMFA and CoMSIA techniques. The developed models will give information about how electrostatic, steric, hydrophobic and hydrogen bonding interactions affect the anti-HIV-1 IN inhibition. These may lead to a better understanding of the molecular mechanisms and structural requirements of IN inhibitors. These may also provide useful information about the design and synthesis of more potent HIV-1 IN inhibitors with predetermined affinity and may show statistically significant results with good predictive ability.

2. Computational details

2.1. Data set for analysis

Anti-HIV IN inhibitory activity data reported by Brzozowski et al. [24,25] was used in this study. The IC_{50} (μM) values were taken in molar (M) range and converted to pIC_{50} according to the formula.

$$pIC_{50} = -\log IC_{50} \tag{1}$$

Activity (IC_{50}) values of 11 molecules were reported $>100 \, \mu M$, however these molecules were selected for study to understand structural features which made them inactive. For sake of simplicity, all of them were considered IC_{50} equal to $100 \, \mu M$. During the model development and validation, it was observed that molecule 14 was outlier. Therefore, this molecule was removed from data set. Total set of 41 molecules was divided in to training set (31 molecules) and test set (10 molecules) randomly. The structures of both the training and test set molecules are shown in

Table 1. This data set was used to construct 3D-QSAR (CoMFA and CoMSIA) model and to analyze their physico-chemical properties.

2.2. Docking studies

To locate the appropriate binding orientations and conformations of these series of HIV-1 IN inhibitors, docking was performed by using FlexX program in SYBYL7.1 software package. FlexX is extensively used [23,27,32,33] fast automated docking program that is based on incremental construction algorithm and considers only ligand flexibility [34,35]. The HIV-1 IN protein (1QS4) was used for study. The missing residues at the position 141–144 were incorporated from monomer B in crystal structure of 1BIS.pdb after superposition of the backbone of crystal structure of 1QS4.pdb and 1BIS.pdb [26]. After building the molecules, HIV-1 IN protein (1QS4) was prepared by using Protein Preparation tool in SYBYL Biopolymer module. The ligand within the active site (5-CITEP) and all the water molecules were removed while metal ion (Mg^{2+}) was allowed to remain with a charge of $+2$. The active site was defined to include all atoms within 6.5 Å radius of the Mg^{2+}. Formal charges and verbosity $= 2$ were assigned and 30 top ranked docking poses were saved for each docking run. All the molecules were built using SYBYL7.1 molecular modeling package installed on a Silicon Graphics Fuel Work station running IRIX 6.5 [36]. Tripos force field, Gasteiger Huckel partial atomic charges [37] and Powell's conjugate gradient method were used for minimization of all molecules with 0.05 kcal/mol energy gradient convergence criterion [38]. The X-ray crystal structure of molecules of these series has not been determined till now. Best docked conformation of highly active molecule 32 was used for alignment of all the molecules in the series. The rest of the molecules were built by changing required substitution on best docked conformation and energy minimized as stated previously. These molecules were then used to construct 3D-QSAR (CoMFA and CoMSIA) model. Autodock4 program was also used for docking of molecule 32 for hypothetical validation of FlexX docking results. AutoDock tools (ADT) were used to prepare the molecule 32, protein (deleting all water molecules, adding polar hydrogen's and loading Kollman United Atoms charges) and also to perform docking calculations. A grid box with spacing 0.375 and dimensions $52 \times 66 \times 46$ points was constructed around the binding site, based on the location of the co-crystallized ligand. All bond rotations and torsions for the ligand were automatically set in the ADT. The Lamarckian genetic algorithm (LGA) was employed and the docking runs were set to 50 [39].

2.3. Alignment of molecules

In the 3D-QSAR studies, alignment and bioactive conformation selection are two important factors for obtaining robust and meaningful models. Best docked conformation of molecule 32 was taken as template for alignment and rest of molecules aligned to it using FIT ATOM method with the common substructure * in molecule 32 (Fig. 1). The resulting alignment is shown in Fig. 2.

2.4. CoMFA interaction energy calculations

CoMFA calculates steric fields using the Lennarde–Jones potentials, and electrostatic fields using the Coulomb potential. Both the steric and the electrostatic fields were calculated for each molecule using a sp^3 carbon atom with charge $+1$ served as probe atom and energy cutoff was set to 30 kcal/mol. A distance dependent dielectric constant 1.00 was used and grid spacing was set to 2 Å. The pattern of 3D cubic lattice generated automatically by using CoMFA routine, extended at least 4.0 Å beyond the volumes of all investigation molecules along the axes.

Table 1
Data set structures (molecules 1–43) and pIC$_{50}$ (actual and predicted) used for 3D-QSAR analysis (CoMFA and CoMSIA).
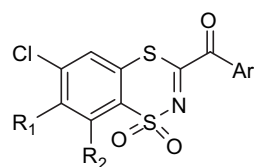
1-17

Molecule	R$_1$	R$_2$	Ar	Actual pIC$_{50}$	Pred. pIC$_{50}$ (CoMFA)	Pred. pIC$_{50}$[a] (CoMSIA)
1	Me	H	Ph	4.92	4.61	4.64
2	Me	H	4-MeO–Ph	4.52	4.63	4.65
3	Me	H	4-Br–Ph	4.52	4.62	4.66
4[b]	Me	H	4-Cl–Ph	4.52	4.62	4.65
5[b]	Me	H	4-F–Ph	4.24	4.59	4.64
6[e]	Me	H	4-O$_2$N–Ph	NT	–	–
7	Me	H	3-O$_2$N–Ph	4.24	4.47	4.4
8[b]	Me	H	3,4-Cl–Ph	4.82	4.56	4.65
9	Me	H	4-Ph–Ph	4.96	4.94	4.88
10	Me	H	2-Naphthyl	4.55	4.55	4.65
11	H	Me	4-Cl–Ph	4.57	4.67	4.65
12[b]	H	Me	3,4-di-Cl Ph	4.77	4.62	4.65
13	H	Me	2-Naphthyl	4.66	4.59	4.65
14[c]	H	H	4-Cl–Ph	4.00	4.68	4.67
15	H	H	3-O$_2$N Ph	4.60	4.55	4.43
16	H	H	3,4-Cl–Ph	4.72	4.63	4.68
17[b]	H	H	2-Naphthyl	5.10	4.62	4.68
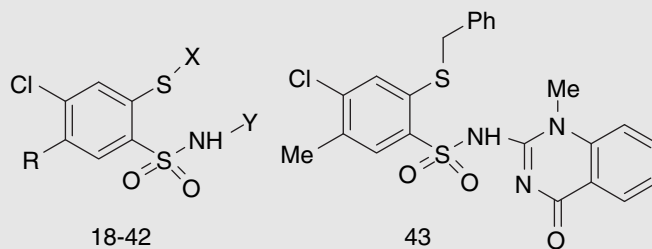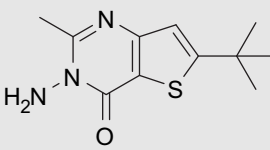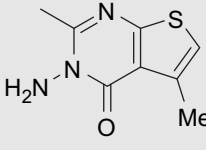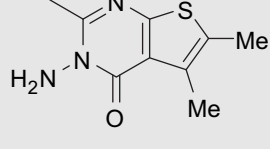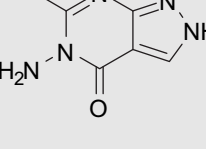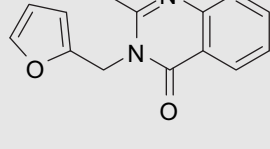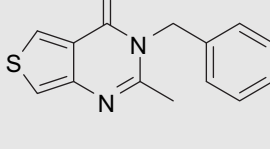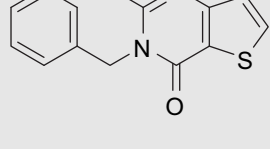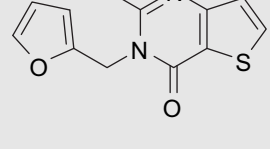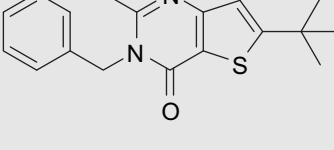
18-42   43

Mole	R	X	Y	Actual pIC$_{50}$	Pred. pIC$_{50}$ (CoMFA)	Pred. pIC$_{50}$[a] (CoMSIA)
18[d]	Me	H		4.00	3.98	4.01
19[d]	Me	H		4.00	3.98	4.10
20[b]	Me	H		4.16	3.92	3.97
21	Me	H		4.18	4.13	4.09

Table 1 (*continued*)

Mole	R	X	Y	Actual pIC$_{50}$	Pred. pIC$_{50}$ (CoMFA)	Pred. pIC$_{50}$[a] (CoMSIA)
22[b]	Me	H		4.22	4.52	4.85
23[d]	Me	H		4.00	4.03	4.01
24	Me	H		4.25	4.32	4.28
25[d]	Me	H		4.00	4.04	4.00
26	Me	H		4.11	4.20	4.16
27	Me	H		4.10	4.11	4.06
28	Me	H		4.28	4.18	4.18
29	Me	H		4.21	4.2	4.21
30[b]	Me	H		4.66	4.79	4.93

(*continued on next page*)

Table 1 (continued)
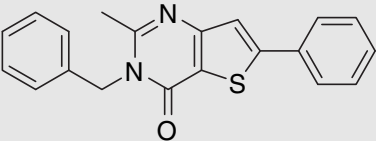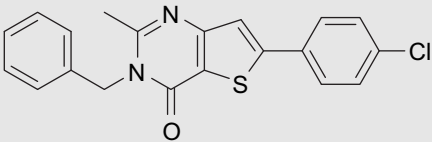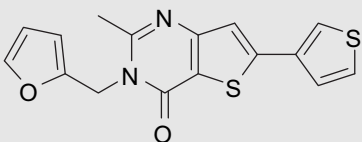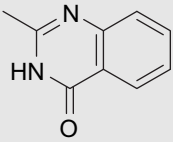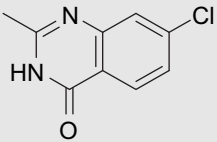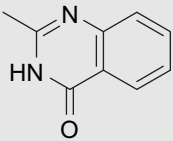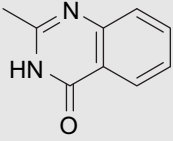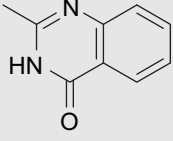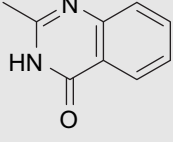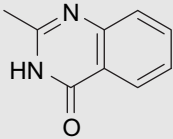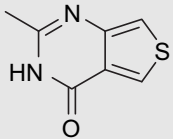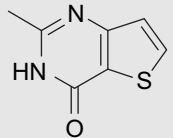
Mole	R	X	Y	Actual pIC$_{50}$	Pred. pIC$_{50}$ (CoMFA)	Pred. pIC$_{50}$[a] (CoMSIA)
31	Me	H		5.05	5.09	5.14
32[b]	Me	H		5.22	5.17	5.23
33	Me	H		5.10	5.09	5.10
34[d]	Me	−CH$_2$Ph		4.00	4.03	4.05
35	Me	−CH$_2$Ph		4.33	4.25	4.27
36[d]	H$_2$NCO	−CH$_2$Ph		4.00	4.01	4.03
37	PhNHCO	−CH$_2$Ph		4.14	4.07	4.03
38[d]	4-MePhNHCO	−CH$_2$Ph		4.00	4.04	4.02
39[d]	4-MeOPhNHCO	−CH$_2$Ph		4.00	4.03	4.02

Table 1 (continued)

Mole	R	X	Y	Actual pIC$_{50}$	Pred. pIC$_{50}$ (CoMFA)	Pred. pIC$_{50}$[a] (CoMSIA)
40	4-ClPhNHCO	–CH$_2$Ph		4.00	4.05	4.02
41[b,d]	Me	–CH$_2$Ph		4.00	3.95	3.98
42[d]	Me	–CH$_2$Ph		4.00	3.99	4.06
43[d]	–	–	–	4.00	3.94	3.99

[a] Predicted pIC$_{50}$ value of best CoMSIA (SH) model.
[b] Test set molecules.
[c] Outlier molecule which was removed from data set.
[d] Molecules have >100 IC$_{50}$ μM considered their IC$_{50}$ equal to 100 μM; remaining molecules included into training set.
[e] NT: molecule 6 activity (IC$_{50}$) data is not given (Ref. [24]) so not included during analysis.

2.5. CoMSIA interaction energy calculations

The five physicochemical properties, comprising steric, electrostatic, hydrogen bond donor and acceptor and hydrophobic interactions were calculated at each lattice intersection of a regularly spaced grid of 2.0 Å. An sp^3 carbon probe atom with radius 1.0 Å and +1 charge, with hydrophobicity value +1, hydrogen bond donor and hydrogen bond acceptor properties +1 was used to calculate steric, electrostatic, hydrophobic, donor and acceptor fields.

2.6. Partial least square (PLS) analysis

PLS method was used to linearly correlate the CoMFA and CoMSIA fields to the inhibitory activity values. The cross-validation analysis was performed using the leave-one-out (LOO) method in which one molecule was removed from the data set and its activity was predicted using the model derived from the rest of the dataset [40]. The optimum number of components was determined by cross-validation method that produced the smallest root mean predictive sum of squared errors, which corresponds to the highest cross-validated r^2_{cv}. To speed up the analysis and to reduce noise, a column filter value σ of 2 kcal/mol was used. Final analysis was performed to calculate conventional r^2 (r^2_{ncv}) using the optimum number of components obtained from the cross-validation analysis. To further assess the robustness and statistical confidence of the derived models, bootstrapping analysis for 100 runs was performed. Bootstrapping involves the generation of many new data sets from original data set and is obtained by randomly choosing samples from the original data set. The statistical calculation is performed on each of these bootstrapping samplings. The difference between the parameters calculated from the original data set and the average of the parameters calculated from the many bootstrapping samplings is a measure of the bias of the original calculations. The all cross-validated results were analyzed by considering the fact that a value of cross-validated r^2_{cv} above 0.3 indicates that probability of chance correlation is less than 5% [41].

2.7. Predictive correlation coefficient (r^2_{pred})

The predictive power of the 3D-QSAR models was determined from a set of 10 molecules that was never used in training set during model development. These molecules were aligned in the same way as those in the training set and their activities were predicted using the model produced by training set. The predictive correlation coefficient (r^2_{pred}) based on the test set molecules, is calculated using Eq. (2):

$$r^2_{pred} = (SD - PRESS)/SD \qquad (2)$$

Where SD is the sum of square deviation between the biological activities of the test set and mean activity of the training set molecules, and PRESS is the sum of squared deviation between predicted and actual activity values for every molecule in test set.

3. Results and discussion

3.1. Docking analysis

The purpose of docking studies is to generate the appropriate binding orientations and conformations of highly active molecule 32. Flexible docking of highly active molecule 32 was carried out in the active site of the HIV-1 IN. Best docked conformation of molecule 32 was selected as most reliable conformation of molecule 32 and used for alignment of 3D-QSAR study. Best docked and most reliable conformation of molecule 32 had been selected on the basis of reported data [10,14,42–50]. The HIV-1 IN active site has two cavities (3P and ST) which are located perpendicular to each other. The ST cavity region begins from Phe139 to Lys159 and
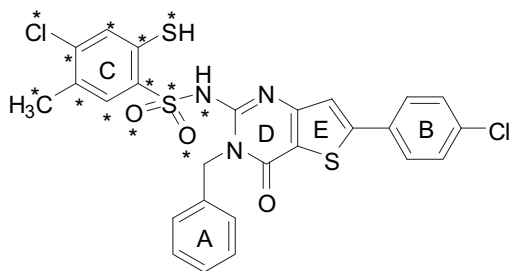
Fig. 1. Highly active molecule 32 and asterisks showed common substructure used for alignment.

interacts with DNA, while 3P cavity comprises of Asp64, Asp116, Asn120, Cys65, Thr66, His67 and Mg^{2+} and has important role in 3′ processing activity. The 3′ processing reaction occurs near metal ion and its presence is also required for activity. The ligands (S1360, BDKA and 5-CITEP) [42] formed coordinate bond to Mg^{2+} which is important for inhibition [10,14,43–45] (3′ processing activity) of HIV-IN. This is already reported that binding to Cys65 and Mg^{2+} chelation are important for inhibition of HIV-1 IN [46]. So these evidences suggest that a ligand might exhibit 3′ processing activity if it binds to above mentioned residues and chelate the Mg^{2+}. The best possible binding pose and most reliable conformation of molecule 32 in active site is illustrated along with X-ray crystal structure of 5-CITEP in Fig. 3. This study reveals that molecule 32 forms hydrogen bonding with Cys65, His67 and Asn155. The docking gesture of the sulfonamide moiety, thiol group and nitrogen of pyrimidine ring formed a coordination ring around the Mg^{2+} which strongly chelate the Mg^{2+} than the carboxyl moieties of Asp64 and Asp116, finally inhibiting the HIV-1 IN activity (Fig. 4). This interaction may be charge–charge interaction between the metal ion and partial or ionic charges of inhibitors. The position of phenyl ring A of molecule 32 found in close proximity of Lys156 and Lys159 which are already proved to be important residues for binding of ligand to the active site and inhibition as well [47–50]. These interactions are in agreement with reported data [10,14,42–50]. This docking analysis revealed that molecule 32 is differently bound to active site than the 5-CITEP, because of structural diversity. Docking studies showed that most of the test and training molecules occupied the same space near the Mg^{2+} in vicinity of active site of HIV-1 IN and have a common binding mode with Cys65, Asn155 and Mg^{2+}. This implies that these three interactions are important to binding of these inhibitors to the active site. The wide range of binding interactions formed by the molecules as well as lack of correlation between FlexX score and inhibitory activities may be due to the structural diversities among the
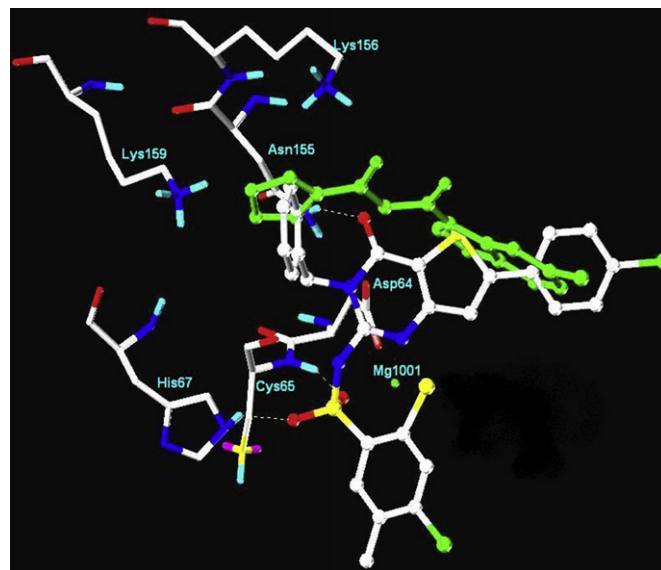

Fig. 3. Docking structure of highly active molecule 32 bound in active site of HIV-1 IN and green ball represents coordinated Mg^{+2} ion. The green capped sticks shows X-ray crystallized structure of 5-CITEP (using FlexX program). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

molecules (Table 3 in Supplementary data for FlexX score with pIC_{50}, and docking position and H-bonding interaction).

In addition, for molecule 32, docking was also performed by Autodock4 program for hypothetical validation of FlexX docking results. It was found that molecule 32 docked in similar manner as found in FlexX docking program in term of location (near to Mg^{2+}), binding mode (Cys65, His67, Asn155). As described previously, best docked conformation of molecule 32 was used as template for 3D-QSAR studies, we superimposed the selected best docked conformations of molecule 32 generated by Autodock4 and FlexX docking programs to see if both the programs gave the similar conformation. The RMSD of superimposed conformations was found 0.0 that means both conformations superimposed completely one above the other. This concordance in Autodock4 and FlexX results further strengthen our selection of conformation of molecule 32 used for alignment for 3D-QSAR studies (Fig. 5). (In Supplementary data Fig. 12 displays FlexX docked conformation and Autodock4 conformation of molecule 32 to the active site of HIV-1 IN).
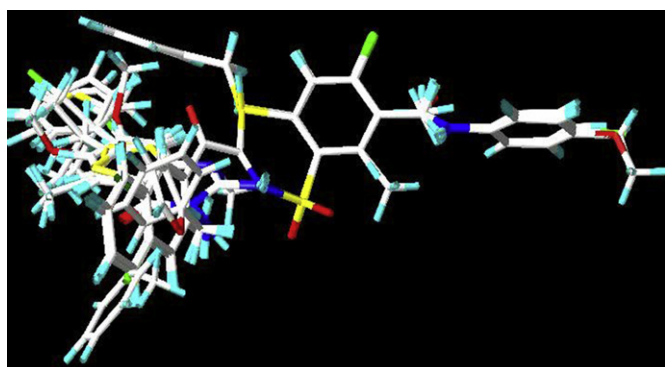

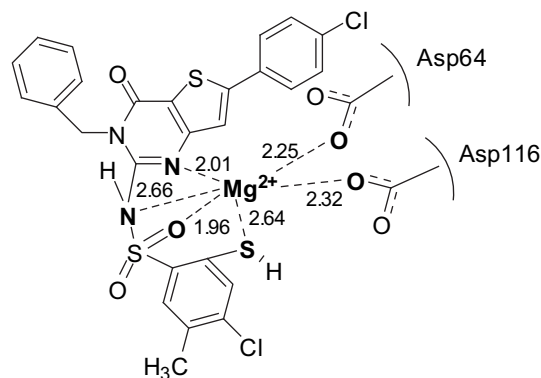**Fig. 2.** Alignment of molecules used in CoMFA and CoMSIA.

Fig. 4. 2D representation of interactions between magnesium ion and molecule 32 obtained from docking to HIV-IN active site.
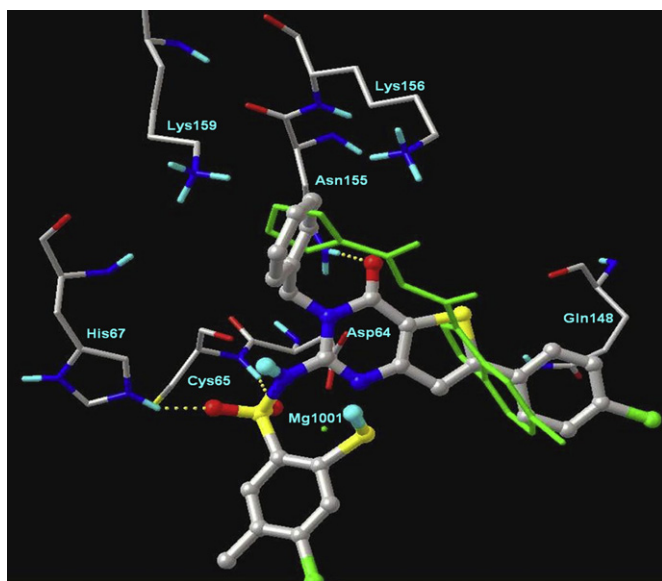
Fig. 5. Docking structure of highly active molecule 32 bound in active site of HIV-1 IN and green ball represents coordinated Mg^{+2} ion. The green capped sticks shows X-ray crystallized structure of 5-CITEP (using Autodock4 program).(For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. CoMFA analysis

The CoMFA and CoMSIA studies were performed based on the alignment of best docked conformation of molecule 32 which was obtained from docking. The data set consisting of 41 molecules was divided randomly into test set of 10 and training set of 31 molecules. The molecules in the training and test set were selected in such a way so as to maintain the maximum diversity of the structure and cover all activity range of these series. During the CoMFA analysis, molecule 14 was found outlier. Here, docking study also supported that molecule 14 is outlier because it was bound away from the Mg^{2+} in the vicinity of HIV-1 IN active site (described below). The final model was generated when molecule 14 is removed from dataset during the model generation by using test set (10 molecules) and training set (31 molecules). The statistical results of CoMFA PLS

analysis is presented in Table 2. The constructed model is robust with cross validated r^2_{cv} of 0.728 having 5 optimum numbers of components which indicated a good predictive capacity of the model. The non-cross validated PLS analysis resulted correlation coefficient (r^2_{ncv}) of 0.934, $F = 70.446$ and estimated standard error of 0.100 with same component used in cross validation calculation. A high r^2_{ncv} showed the self consistency of the model. The goodness of internal prediction is represented by cross validated r^2_{cv} whereas goodness of fit of a QSAR model is represented by r^2_{ncv}. To test the external predictive ability of model, test set of 10 molecules was used, which were not included in model generation. The predictive correlation coefficient (r^2_{pred}) of 0.708 indicated good external predictive ability of CoMFA model. It can be seen clearly from Table 1 that the predicted pIC_{50} values of the test set molecules are in tolerable residual range with the experimental data (actual pIC_{50}). The higher r^2_{bs} value of 0.966 obtained after 100 runs of bootstrapping supports this analysis. In addition to this low value of standard error of estimates (0.100), PRESS (0.634) also supports the significance of the developed model. The steric and electrostatic fields' contribution in CoMFA is 74% and 26%, respectively which implies that the contribution of steric part is predominant for interaction of these inhibitors to HIV-1 IN.

The scatter plot for the actual pIC_{50} versus predicted pIC_{50} values for test set and training set is shown in Fig. 6. Table 1 shows predicted pIC_{50} values of the molecules under the CoMFA model development.

3.3. CoMSIA analysis

CoMSIA uses three more physiochemical properties- hydrophobic, hydrogen-bond donor and acceptor which CoMFA does not use. Thus different combinations of the three properties with steric and electrostatic properties may result in different models. So, in this study many models were built by varying these five physicochemical properties (shown in Table 2). Model SH was considered best CoMSIA model and was used to generate contour maps. With this model cross validated r^2_{cv} of 0.794 of 5 components, non-cross validated r^2_{ncv} of 0.928, F-test of 67.34 and SEE of 0.104 were obtained. The contributions of steric and hydrophobic properties are 46% and 54%, respectively. The results indicate that hydrophobic property has more contribution than the steric property in CoMSIA model development but CoMFA showed steric property. In

Table 2
PLS summery results for CoMFA and CoMSIA models.

| | CoMFA | CoMSIA | | | | | | | | |
		SE	SH	SEH	SHD	SHA	SEHD	SEHA	SHDA	SEHDA
NOC	5	6	5	6	6	6	6	6	6	6
r^2_{cv}	0.728	0.615	0.794	0.703	0.777	0.78	0.69	0.732	0.735	0.698
r^2_{bs}	0.966	0.969	0.946	0.972	0.971	0.976	0.964	0.972	0.966	0.965
SD	0.2	0.016	0.019	0.01	0.015	0.01	0.016	0.012	0.013	0.014
r^2_{nCV}	0.934	0.922	0.928	0.938	0.932	0.93	0.932	0.935	0.928	0.928
F-test	70.45	47.11	67.34	60.80	54.45	52.36	54.65	57.60	51.6	51.83
SEE	0.1	0.111	0.104	0.098	0.103	0.105	0.103	0.101	0.106	0.106
r^2_{pred}	0.708	0.6	0.59	0.55	0.56	0.58	0.48	0.55	0.57	0.49
PRESS	0.634	0.864	0.890	0.974	0.965	0.902	1.127	0.973	0.930	1.109
Field contributions (%)										
S	74	56	46	34	37	38	28	32	33	26
E	26	44	–	26	–	–	20	21	–	16
H	–	–	54	40	45	45	35	37	39	32
D	–	–	–	–	18	–	17	–	15	16
A	–	–	–	–	–	17	–	10	13	10

PRESS = predictive residual sum of square for the training set; r^2_{cv} = cross-validated correlation coefficient by PLS LOO method; NOC = optimum number of component as determined by PLS LOO cross-validation study; SEE = standard error of estimate; r^2_{ncv} = conventional correlation coefficient (non-cross validation); r^2_{bs} = correlation coefficient after 100 runs of bootstrapping; $S.D_{bs}$ = standard deviation from 100 runs of bootstrapping; r^2_{pred} = predictive correlation coefficient, S for steric, E for electrostatics, H for hydrophobic, D for hydrogen bond donor, A for hydrogen bond acceptor.
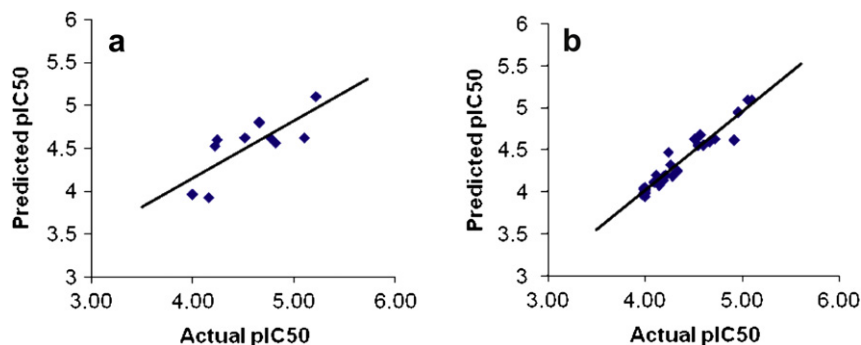
Fig. 6. Scatter plot of actual pIC$_{50}$ versus predicted pIC$_{50}$ of (a) test set and (b) training set for CoMFA model.

addition, CoMSIA analysis reveals that the contribution of steric property is also important, but hydrogen bond donor and acceptor have very less contribution. The decrease in steric property in the CoMSIA model can be explained by the introduction of hydrophobic property in CoMSIA model and also by the fact that in some cases the steric property is somewhat correlated to hydrophobic property. So a conclusion can be drawn that steric and hydrophobic properties play important role in building the model. Most of models (SE, SH, SHE, SHD, SHA, SEHD, SEHA, SHDA, SEHDA) showed that greatest field contribution is hydrophobic field, followed by steric and very low contribution of hydrogen bonding. The model developed by CoMSIA steric and electrostatic properties gave a poor cross validation than the CoMFA model.

The scatter plot for the actual pIC$_{50}$ versus predicted pIC$_{50}$ values for test set and training set are shown in Fig. 7. Table 1 also shows predicted pIC$_{50}$ values of the molecules under the CoMSIA analysis.

3.4. External data set evaluation

In order to establish further the robustness of the CoMFA model, 14 novel benzodithiazine inhibitors of HIV-1 IN recently reported by Brzozowski et al. were evaluated [51]. The actual and predicted pIC$_{50}$ values with structures for these external dataset of 14 molecules are shown in Table 4 in Supplementary data. The CoMFA model showed good predictivity with low residuals for these 14 molecules, which further support reliability of our CoMFA model.

3.5. Contour maps analysis

Table 2 shows that the CoMFA steric and electrostatic properties have varying contributions, accounting for 74% and 26%, respectively, which suggest that both properties are critical in explaining the variations in inhibition potency of these molecules. But the steric property is more important. The CoMFA contour maps of steric (yellow and green) and electrostatic (red and blue) properties are shown in Figs. 8 and 9, respectively and with molecule 32 along with active site of HIV-1 IN. In CoMFA steric contour, green color indicates sterically favored regions where addition of bulky group increases the activity while yellow color indicates sterically disfavored regions where addition of bulky group decreases the activity.

The CoMFA PLS steric contour maps for the 3′-processing inhibitory activity, are shown in Fig. 8. A large green contour is covering the phenyl ring B in molecule 32, which is surrounded by the Gln62, Leu63, Ile141, Pro142, Thr115 – these residues form moderately lipophilic binding pocket. Molecules 31, 32 and 33 which contain hydrophobic ring such as phenyl, *p*-chloro-phenyl and thiophenyl may have good hydrophobic interaction with these residues. This is a possible reason why these molecules have higher potency. Most active molecule 32 has *p*-chlorophenyl moiety that contributes to high hydrophobic interaction with these residues and improved the binding affinity. This is also reflected in molecule 9, which has bulky substituents (–Ph–Ph) and improved activity. Most of molecules in this series are not covered by this green contour because of lack of these rings in their structure so they have less potency. A big yellow contour is covering the ring E; large substitutions in this region may reduce activity. It may explain why molecules 19, 22, 23 and 24 are less potent than molecule 32. This is also reflected in molecules 10, 13 and 17, which have bulky substituents (2-naphthyl) at this position and have low potency. This yellow contour found near the key residues Asp64 and Asp116, which were reported to be involved catalytic process of HIV-1 IN [19,52]. Three small yellow contours are covering the phenyl ring A. This sterically disfavored area is blocked by the Asn155, Lys156, and Lys159.
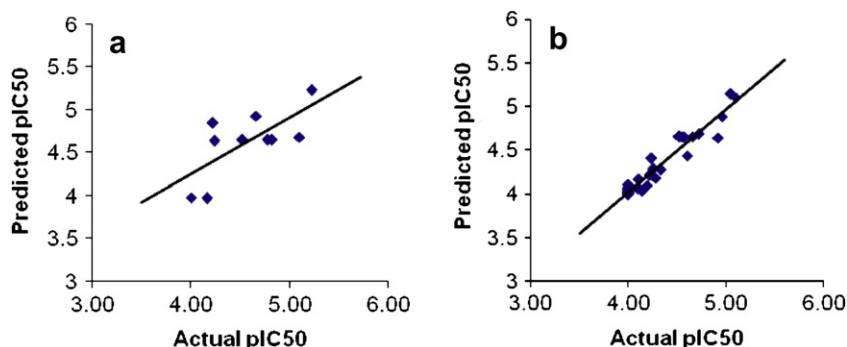

Fig. 7. Scatter plot of actual pIC$_{50}$ versus predicted pIC$_{50}$ of (a) test set and (b) training set for CoMSIA model.
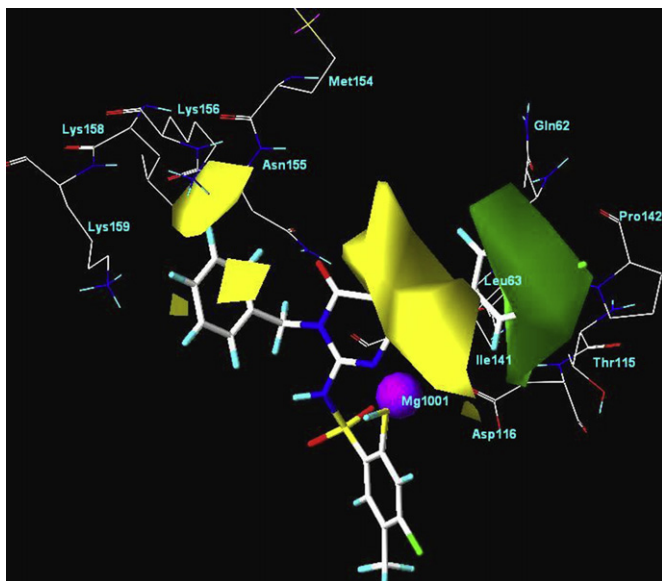
Fig. 8. CoMFA steric contour maps: green contour represent sterically favored regions; yellow contour represent sterically disfavored regions. The balls and sticks show molecule 32 in to the active site of HIV-1 IN. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

so that the negative substituents should strengthen the binding of the inhibitors.

A total of nine CoMSIA models were generated using either single or combined fields. Most of the models which have hydrophobic and steric fields showed a high cross-validated r^2_{cv} value, which indicates the importance of hydrophobic and steric properties for these series of molecules. The SH model is considered best CoMSIA model on the basis of high cross validated r^2_{cv} value that was used for contour maps generation. The docking studies showed that hydrogen bonding is also important for interaction of ligand to active site of HIV-1 IN, but were not considered during analysis because of relatively less contribution of hydrogen bonding as compared to steric and hydrophobic properties and low value of cross validated r^2_{cv} as compare to SH model (Table 2).

Fig. 10 shows the hydrophobic contour maps. Yellow contour refers to areas where the hydrophobic substituents are favorable to improve inhibitor activities, while the white contour indicates the area where hydrophilic substitutions are favorable to increase inhibitory activity. A large white contour was covering the ring E and interacts with the same residues as found in CoMFA yellow contour map. So, substitution of hydrophilic group (which have less steric property) may improve the affinity of inhibitor. One small white contour was found near ring A that surrounded Lys156, which has hydrophobic disfavourable region. This region was also found in CoMFA steric contour maps. One small yellow contour was found near the phenyl ring B in molecule 32. This yellow contour is surrounded by the Leu63 and Ile141, which have large contribution in forming of moderately lipophilic binding pocket (as described in CoMFA green contour map). One yellow contour was found near the phenyl ring A, indicating hydrophobic favorable region. One more yellow contour was found near the nitrogen of ring D, so substitution of hydrophobic groups (which have steric property) at this site may improve affinity of inhibitors. Fig. 11 shows the steric maps, which are comparable to CoMFA steric contour maps.

On the other hand, in case of electrostatic contour (Fig. 9), red color indicates regions where the addition of electronegative substituents increases activity, while blue color indicates regions where the addition of electropositive substituents increases activity. A blue contour is found near the phenyl ring A indicating negatively charged substitution in the area is unfavorable. It can explain the fact that activity of molecules 18–25 (–NH$_2$) are lower than the highly active molecule 32. One smaller blue contour on phenyl ring B suggests that electropositive group improves the activity. A red contour map near the sulfonamide moiety and surrounded the Cys65 and His67, both residues act as hydrogen donor
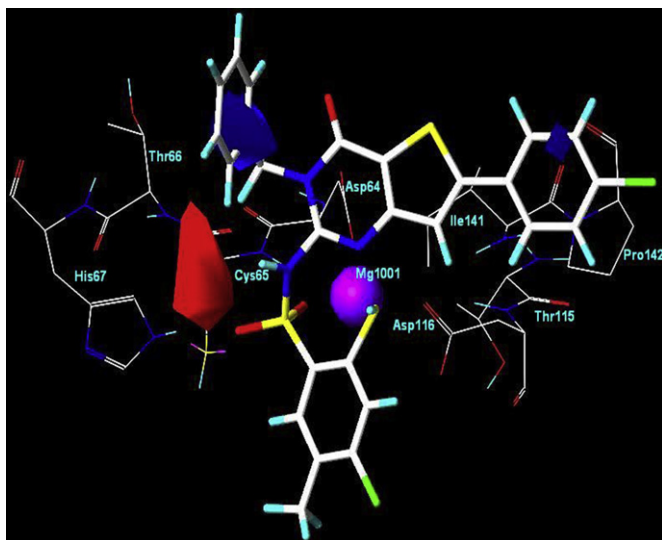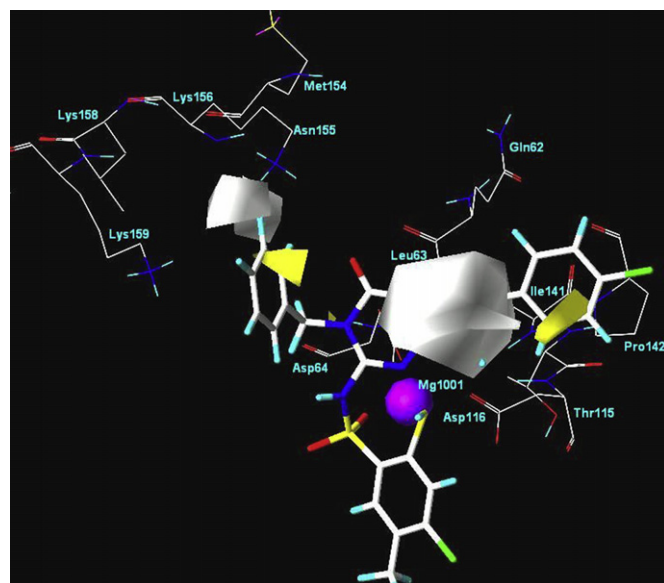

Fig. 9. CoMFA electrostatic contour maps: red contour represent negative potential favored, blue contour represent positive potential favored. The balls and sticks show molecule 32 in to the active site of HIV-1 IN. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 10. CoMSIA hydrophobic contour maps: yellow contour represent hydrophobic substitution favored the activity, white contour represent hydrophobic substitution disfavored the activity. The balls and sticks show molecule 32 in to the active site of HIV-1 IN. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 11. CoMSIA steric contour maps: green contours represent sterically favored regions, yellow contour represent sterically disfavored regions. The balls and sticks show molecule 32 in to the active site of HIV-1 IN. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
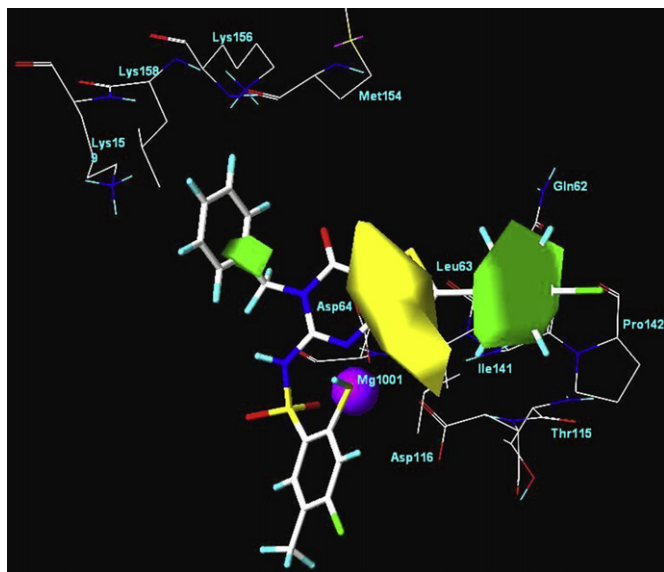
4. Conclusions

In this study, molecular docking has been employed to identify a potential binding mode for these inhibitors at HIV-1 IN active site and best docked conformation of molecule 32 was used as template for alignment. It was found that three interactions (Mg^{2+}, Cys65 and Asn155) are important for these molecules to bind to active site of HIV-1 IN. Two of these interactions are already in agreement with previously published results. However, the present study also indicates that there is another residue (Asn155) which contributes to the binding of these ligands to the HIV-1 IN for 3′ processing activity. But it was found that one molecule 14 had a totally different docking pose than the other molecules. It was interacting away from Mg^{2+} in the active site by residues which are different from the others. As previously mentioned that molecule 14 was found as outlier during model developments. There are several reasons for outlier occurrence in QSAR studies, for example, molecules might be a result of the random experimental error that might be significant when analyzing large data sets. It is also possible that outlier might be acting by a mechanism different from that of the majority of the data set. The docking results showed that molecule 14 was interacting differently with the active site by forming H-bond to His67, Asn155, Lys156 and Lys159 and away from the Mg^{2+} (Fig. 13 in supplementary data). Similarly, docking was performed on other molecules 3 (−Br), 4 (−Cl), 5 (−F) and 11 (R_1 = H and −Cl) that have almost similar structure to molecule 14. This was done to check if these molecules differ in their respective binding modes. The molecules (3, 4, 5, and 11) were interacting near Mg^{2+}, while molecules 3, 4, 5 and 11 formed hydrogen bond to Cys65 and Asn155 in active site of HIV-1 IN, respectively (Shown in Supplementary data Figs. 14–17). Chemically, molecule 14 and other molecules (3, 4, 5 and 11) (Tables 1 and 2) have sufficient diversity in their structures which resulted in different interaction of these molecules to the active site. Final result of these comparison studies showed that molecule 14 did not interact to Cys65 and Mg^{2+} (with which most of the molecules interacted) and therefore it was found as an outlier in 3D-QSAR analysis.

When all the molecules which did not have exact value of activity ($>IC_{50}$) were removed from the data set during model development, a very low value of cross validated r^2_{cv} (<0.5) was obtained. Upon inclusion of these molecules, cross validated r^2_{cv} improved substantially. Therefore, some structural features of these molecules (which make these molecules inactive) are also required for building a statistically significant model. These molecules may also cover some chemical space which may not be covered by the molecules with exact activity value. Therefore, a satisfactory model should be the one which incorporates structural information from both active as well as inactive molecules and is able to discriminate between the both. We have also established predictive CoMFA and CoMSIA 3D-QSAR models for the HIV-1 IN inhibitors. Both models from CoMFA and CoMSIA are satisfactory according to the statistical validation results as well as contour map analysis. The satisfactory CoMFA and CoMSIA were obtained with cross validated r^2_{cv} values of 0.728 and 0.794, respectively, and the non-cross validated PLS analysis with the optimum number of components of 5 having non-cross validated r^2_{ncv} values of 0.934 and 0.928, respectively. All constructed models possessed good internal and external consistency and showed statistical significance and predictive abilities. The CoMFA and CoMSIA results showed that steric and hydrophobic properties are important for binding of these molecules to active site. A combined application of the obtained CoMFA and CoMSIA models may be further employed for the design of new HIV-1 IN inhibitors with improved inhibitory activities.

Appendix. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.ejmech.2009.07.010.

References

[1] M.J. Gait, J. Karn, Trends Biotechnol. 13 (1995) 430–438.
[2] J. d'Angelo, J.F. Mouscadet, D. Desmaële, F. Zouhiri, H. Leh, Pathol. Biol. 49 (2001) 237–246.
[3] M.D. Andrake, A.M. Skalka, J. Biol. Chem. 271 (1996) 19633–19636.
[4] J. Ding, K. Das, H. Moereels, L. Koymans, K. Andries, P.A.J. Janssen, S.H. Hughes, E. Arnold, Nat. Struct. Biol. 2 (1995) 407–415.
[5] A.A. Johnson, C. Marchand, Y. Pommier, Curr. Top. Med. Chem. 4 (2004) 1059–1077.
[6] H. Yuan, A.L. Parrill, Theochem 529 (2000) 273–282.
[7] D.D. Richman, Nature 410 (2001) 995–1001.
[8] Y. Pommier, A.A. Johnson, C. Marchand, Nat. Rev. Drug Discov. 4 (2005) 236–248.
[9] Y. Goldgur, R. Craigie, G.H. Cohen, T. Fujiwara, T. Yoshinaga, T. Fujishita, H. Sugimoto, T. Endo, H. Murai, D.R. Davies, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 13040–13043.
[10] E. Asante-Appiah, A.M. Skalka, J. Biol. Chem. 272 (1997) 16196–16205.
[11] G. Bujacz, J. Alexandratos, A. Wlodawer, G. Merkel, M. Andrake, R.A. Katz, A.M. Skalka, J. Biol. Chem. 272 (1997) 18161–18168.
[12] F. Zouhiri, J.F. Mouscadet, K. Mekouar, D. Desmaele, D. Savoure, H. Leh, F. Subra, M. Le Bret, C. Auclair, J. d'Angelo, J. Med. Chem. 43 (2000) 1533–1540.
[13] M. Ouali, C. Laboulais, H. Leh, D. Gill, D. Desmaele, K. Mekouar, F. Zouhiri, J. d'Angelo, C. Auclair, J.F. Mouscadet, J. Med. Chem. 43 (2000) 1949–1957.
[14] J.A. Grobler, K. Stillmock, B. Hu, M. Witmer, P. Felock, A.S. Espeseth, A. Wolfe, M. Egbertson, M. Bourgeois, J. Melamed, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 6661–6666.
[15] S.P. Lee, J. Xiao, J.R. Knutson, M.S. Lewis, M.K. Han, Biochemistry 36 (1997) 173–180.
[16] R. Zheng, T.M. Jenkins, R. Craigie, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 13659–13664.
[17] A. Engelman, R. Craigie, J. Virol. 66 (1992) 6361.
[18] A.D. Leavitt, L. Shiue, H.E. Varmus, J. Biol. Chem. 268 (1993) 2113–2119.

[19] T.K. Chiu, D.R. Davies, Curr. Top. Med. Chem. 4 (2004) 965–977.
[20] C. Vink, A.M. Oude Groeneger, R.H. Plasterk, Nucleic Acids Res. 21 (1993) 1419–1425.
[21] S. Deswal, N. Roy, Eur. J. Med. Chem. 42 (2007) 463–470.
[22] N. Nunthaboot, S. Tonmunphean, V. Parasuk, P. Wolschann, S. Kokpol, Eur. J. Med. Chem. 41 (2006) 1359–1372.
[23] C.L. Kuo, H. Assefa, S. Kamath, Z. Brzozowski, J. Slawinski, F. Saczewski, J.K. Buolamwini, N. Neamati, J. Med. Chem. 47 (2004) 385–399.
[24] Z. Brzozowski, F. Saczewski, T. Sanchez, C.L. Kuo, M. Gdaniec, N. Neamati, Bioorg. Med. Chem. 12 (2004) 3663–3672.
[25] Z. Brzozowski, J. Slawinski, F. Saczewski, T. Sanchez, N. Neamati, Eur. J. Med. Chem. 43 (2008) 1188–1198.
[26] Y. Goldgur, F. Dyda, A.B. Hickman, T.M. Jenkins, R. Craigie, D.R. Davies, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 9150–9154.
[27] W. Yuan, L.B. Luan, Y.N. Li, Chin. J. Chem. 25 (2007) 453–460.
[28] J.H. Lee, N.S. Kang, S.E. Yoo, Bioorg. Med. Chem. Lett. 18 (2008) 2479–2490.
[29] P. Yi, X. Fang, M. Qiu, Eur. J. Med. Chem. 43 (2008) 925–938.
[30] R. Hu, F. Barbault, M. Delamar, R. Zhang, Bioorg. Med. Chem. 17 (2009) 2400–2409.
[31] E. Cichero, S. Cesarini, P. Fossa, A. Spallarossa, L. Mosti, Eur. J. Med. Chem. 44 (2009) 2059–2070.
[32] C.B. Xue, L. Zhang, W.C. Luo, X.Y. Xie, L. Jiang, T. Xiao, Bioorg. Med. Chem. 15 (2007) 2006–2015.
[33] S.G. Kini, A.R. Bhat, B. Bryant, J.S. Williamson, F.E. Dayan, Eur. J. Med. Chem. 44 (2009) 492–500.
[34] B. Kramer, M. Rarey, T. Lengauer, Protein: Struct. Funct. Genet. 37 (1999) 228–241.
[35] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, J. Mol. Biol. 261 (1996) 470–489.

[36] SYBYL7.1, Tripos Inc., St. louis, MO., 63144, USA, 2005.
[37] J. Gasteiger, M. Marsili, Tetrahedron 36 (1980) 3219–3228.
[38] M.J.D. Powell, Pathol. Biol. 12 (1977) 241–254.
[39] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, J. Comput. Chem. 19 (1998) 1639–1662.
[40] B.L. Podlogar, D.M. Ferguson, Drug Des. Discov. 17 (2000) 4–12.
[41] M. Clark, R.D. Cramer Iii, D.M. Jones, D.E. Patterson, P.E. Simeroth, Tetrahedron Comput. Methodol. 3 (1990) 47–59.
[42] R. Dayam, N. Neamati, Bioorg. Med. Chem. 12 (2004) 6371–6381.
[43] C. Marchand, A.A. Johnson, R.G. Karki, G.C.G. Pais, X. Zhang, K. Cowansage, T.A. Patel, M.C. Nicklaus, T.R. Burke, Y. Pommier, Mol. Pharmacol. 64 (2003) 600–609.
[44] J. Kulkosky, K.S. Jones, R.A. Katz, J.P. Mack, A.M. Skalka, Mol. Cell. Biol. 12 (1992) 2331–2338.
[45] V. Ellison, P.O. Brown, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 7316–7320.
[46] N. Neamati, Z. Lin, R.G. Karki, A. Orr, K. Cowansage, D. Strumberg, G.C.G. Pais, J.H. Voigt, M.C. Nicklaus, H.E. Winslow, J. Med. Chem. 45 (2002) 5661–5670.
[47] M. Sechi, M. Derudas, R. Dallocchio, A. Dessi, A. Bacchi, L. Sannia, F. Carta, M. Palomba, O. Ragab, C. Chan, J. Med. Chem. 47 (2004) 5298–5310.
[48] R. Dayam, T. Sanchez, O. Clement, R. Shoemaker, S. Sei, N. Neamati, J. Med. Chem. 48 (2005) 111–120.
[49] Y.Q. Long, X.H. Jiang, R. Dayam, T. Sanchez, R. Shoemaker, S. Sei, N. Neamati, J. Med. Chem. 47 (2004) 2561–2573.
[50] C.A. Sotriffer, H. Ni, J.A. McCammon, J. Med. Chem. 43 (2000) 4109–4117.
[51] Z. Brzozowski, F. Saczewski, J. Slawinski, T. Sanchez, N. Neamati, Eur. J. Med. Chem. 44 (2009) 190–196.
[52] R. Craigie, J. Biol. Chem. 276 (2001) 23213–23216.