

Original article

Three-dimensional quantitative structure–activity relationship studies on diverse structural classes of HIV-1 integrase inhibitors using CoMFA and CoMSIA

Nadtanet Nunthaboot^a, Somsak Tonmunpheap^a, Vudhichai Parasuk^a,
Peter Wolschann^{b,*}, Sirirat Kokpol^{a,**}

^a Department of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand

^b Institute for Theoretical Chemistry, University of Vienna, A-1090 Vienna, Austria

Received 16 August 2005; received in revised form 10 March 2006; accepted 1 June 2006

Available online 26 September 2006

Abstract

Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA), three-dimensional quantitative structure–activity relationship (3D-QSAR) techniques, were applied to a set of 89 HIV-1 integrase (IN) inhibitors (training set = 61, test set = 28), belonging to 11 structurally different classes. The biological data for 3' processing mechanism were used. For CoMFA calculations, three different fitting methods for alignment process were investigated. The best CoMFA model yielded the cross-validated r^2 (r^2_{cv}) = 0.698 and the non-cross-validated r^2 (r^2) = 0.947. The derived model indicated the importance of steric (60.8%) as well as electrostatic (39.2%) contributions. For CoMSIA calculations, different combinations of the fields were tested. The best CoMSIA model gave r^2_{cv} = 0.724 and r^2 = 0.864. This model showed that steric (30.3%), hydrogen bond donor (43.4%) and hydrogen bond acceptor (26.3%) properties played major roles in HIV-1 IN inhibition. The mapping of hydrogen bond interaction fields with the HIV-1 IN active site gave details on hydrogen bond forming between ligands and enzyme. These obtained results agree well with the experimental observations that there should be hydrogen bond interactions between ligands and Glu152, Lys156 and Lys159 residues. The results not only lead to a better understanding of structural requirements of HIV-1 IN inhibitors but also can help in the design of new IN inhibitors.

© 2006 Elsevier Masson SAS. All rights reserved.

Keywords: HIV-1 integrase inhibition; 3D-QSAR; CoMFA; CoMSIA

1. Introduction

Acquired immune deficiency syndrome (AIDS) is caused by human immunodeficiency virus (HIV). Three essential enzymes, reverse transcriptase (RT), protease (PR) and integrase (IN), are required in the HIV life cycle. Although commercial drugs inhibiting RT and PR have been developed, their efficiency is limited due to side effects and drug resistance. Therefore, it is necessary to search for other antiviral targets. This makes the HIV-1 IN to

be an attractive target for new developments in anti-AIDS therapy. Drugs targeted to IN would be a valuable complement to RT and PR inhibitors. Combination therapy of RT and PR drugs has been proved to be effective in reducing viral load and HIV-1 mortality and morbidity [1,2]. When used in combination, HIV-1 IN inhibitors are expected to improve the outcome of therapy. At the present time, no drug has been developed inhibiting HIV-1 IN but there are some IN inhibitors e.g. S-1360 and L-870,810 under clinical trials [3].

HIV-1 IN mediates an integration of a DNA copy of viral genome into host chromosome via a two-step mechanism [4]. The first step is 3' processing, where IN removes two nucleotides from each 3' end of the viral DNA. The next step is strand transfer, where IN joins the previously processed 3'

* Corresponding author. Tel.: +43 1 4277 52772; fax: +43 1 4277 9527.

** Corresponding author. Tel.: +662 2187583; fax: +662 2187598.

E-mail addresses: karl.peter.wolschann@univie.ac.at (P. Wolschann), sirirat@atc.atccu.chula.ac.th, siriratkokpol@gmail.com (S. Kokpol).

ends to 5' ends of the strand of the target DNA. HIV-1 IN is composed of three domains. The amino-terminal domain (residues 1–50), a zinc-binding domain, promotes enzyme multimerization and enhances the catalytic activity [5]. The central core domain (residues 50–212) incorporates a catalytic triad of amino acid residues, Asp64, Asp116, and Glu152 that are conserved among all retroviral INs [6]. This domain contains the active site for chemical catalysis. The last domain, carboxyl-terminal domain (residues 212–288), is nonspecific in function but has a strong DNA binding affinity [7].

Three-dimensional quantitative structure–activity relationship (3D-QSAR) techniques, such as comparative molecular field analysis (CoMFA) [8] and comparative molecular similarity indices analysis (CoMSIA) [9], are valuable tools for discovering new drugs. Over the last few years, some QSAR techniques were applied to several classes of HIV-1 IN inhibitors [10–18], e.g. coumarins, quinones, hydrazides, dioxepinones, salicylpyrazolinones, catechol and non-catechol containing compounds and mercaptobenzenesulfonamides. For calculations that included either one or a few classes of HIV-1 IN inhibitors [10–12,16–18], a single 3D-QSAR model could be achieved. However, in case of various classes of IN inhibitors [13–15], a single QSAR model describing the activities of all structural diversity inhibitors could not be generated and all the compounds must be clustered to obtain acceptable models. This means that in order to predict the activities of compounds, at least 2 equations are required because the prediction depends on the structures of ligands. Therefore, it becomes fascinating for us to develop a single 3D-QSAR equation explaining HIV-1 IN inhibitor activities without clustering of compounds. As a result, biological activities of any compound could be predicted by using this single equation.

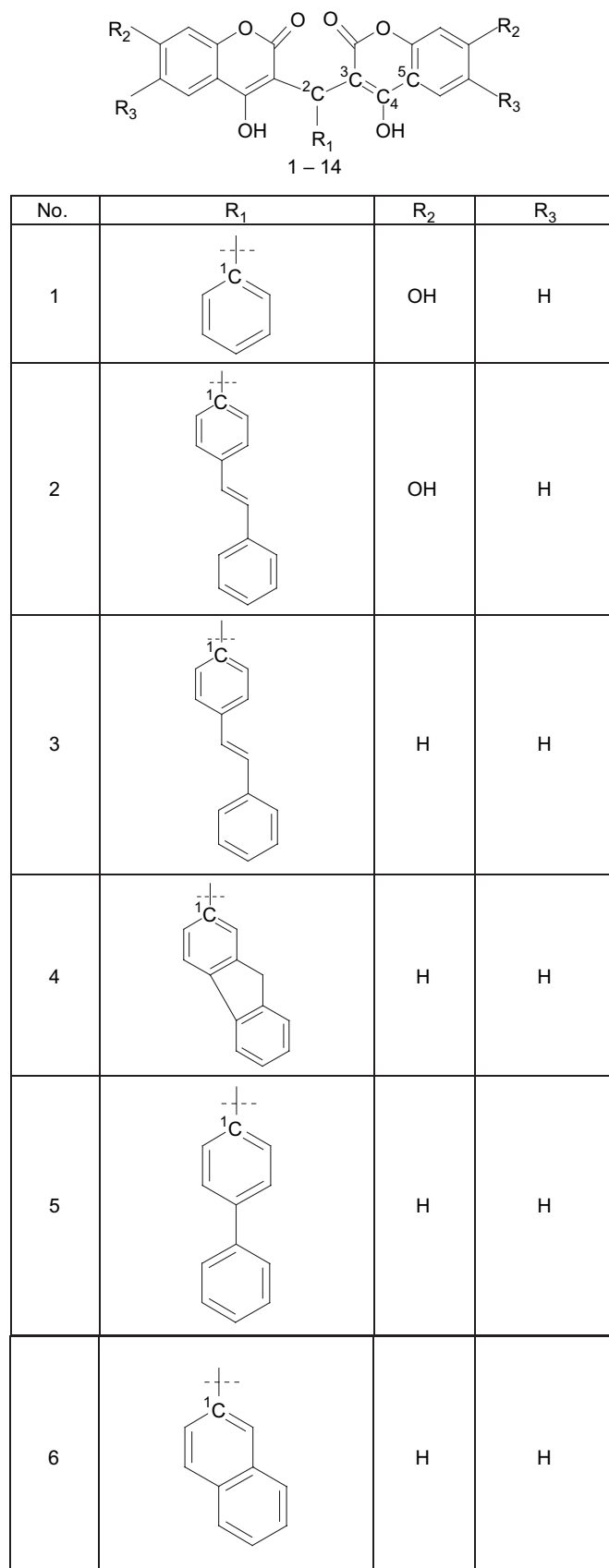
The major objectives of this study are to construct single 3D-QSAR models for the whole variety of 11 structural classes of HIV-1 IN inhibitors and to examine the structure–activity relationship for the inhibition of these inhibitors using CoMFA and CoMSIA techniques. The obtained models will give some insight into how steric, electrostatic, hydrophobic and hydrogen bonding interactions influence anti-HIV-1 IN inhibition. These could lead to a better understanding of the molecular mechanisms and structural requirements of IN inhibitors.

2. Materials and methods

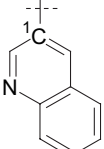
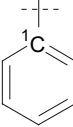
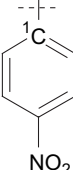
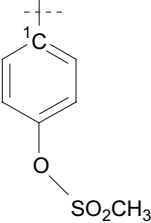
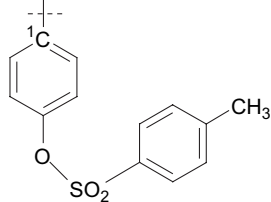
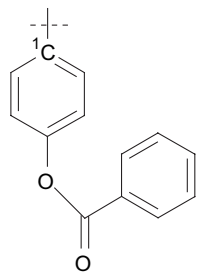
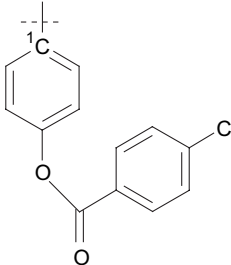
2.1. Biological data

Biological data against the 3' processing mechanism were taken from literature [19–31] and were separated randomly into two subsets; 61 and 28 compounds for training and test sets, respectively. Although the biological activities of compounds were taken from different sources, most of them (more than 70%) were evaluated from the same laboratory. Likewise to the previous QSAR studies [13–15], all compounds were combined together based on the assumption that methods and conditions of activity testing were similar.

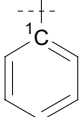
All molecular structures and their aligned atoms are depicted in Scheme 1 (training set) and Scheme 2 (test set).

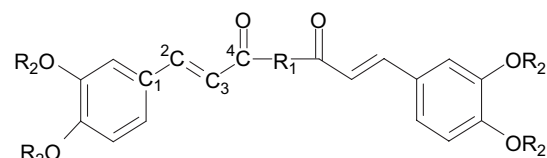


Scheme 1. Structures of compounds used in the training set. The labeled numbers correspond to atoms C1–C5 of 5CITEP (template) used in superposition process.

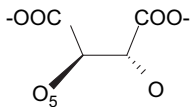
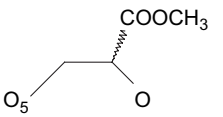
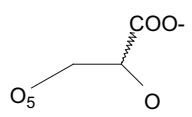
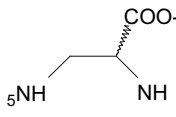
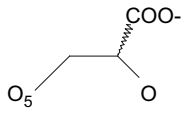
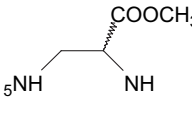
No.	R ₁	R ₂	R ₃
7		H	H
8		H	H
9		H	H
10		H	H
11		H	H
12		H	H
13		H	H

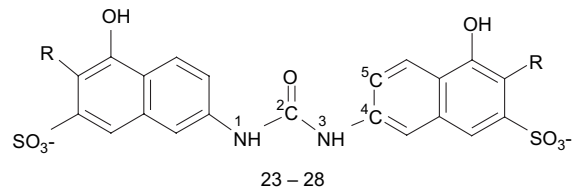
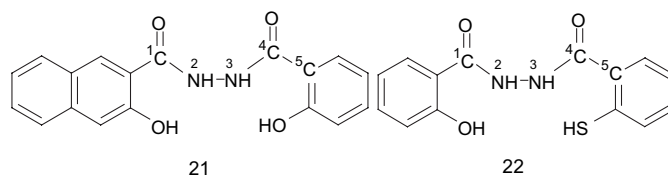
Scheme. 1 (continued).

No.	R ₁	R ₂	R ₃
14		OH	(CH ₂) ₅ CH ₃



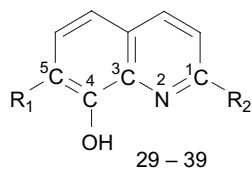
15 – 20

No.	R ₁	R ₂
15		CH ₂ CH ₃
16		CH ₂ CH ₃
17		CH ₂ CH ₃
18		CH ₂ CH ₃
19		H
20		CH ₂ CH ₃



Scheme. 1 (continued).

No.	R
23	H
24	
25	
26	
27	
28	

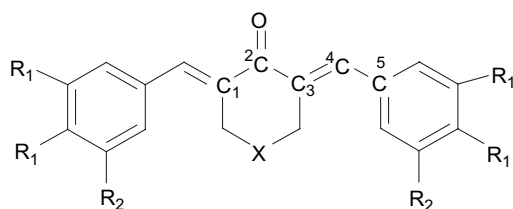


No.	R ₁	R ₂
29	COO ⁻	
30	COO ⁻	
31	COO ⁻	
32	COONa	
33	COO ⁻	
34	COO ⁻	

Scheme. 1 (continued).

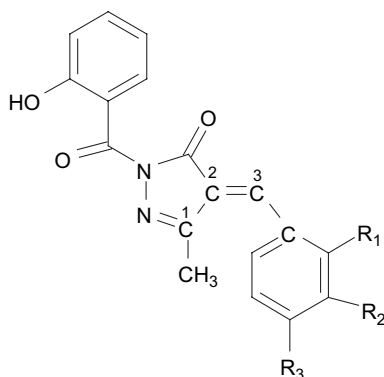
No.	R	
35	COO ⁻	
36	COO ⁻	
37	COO ⁻	
38	COO ⁻	
39	COO ⁻	
40		
41		
42		
43		
44		

Scheme. 1 (continued).



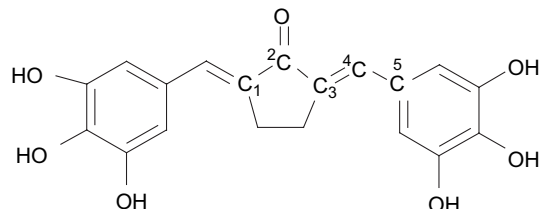
45 – 50, 52

No.	R ₁	R ₂	X
45	Cl	H	O
46	OH	H	NCH ₃
47	OH	H	NCH ₂ CH ₃
48	OH	H	NCH ₂ Ph
49	OH	OH	O
50	OH	OH	S
52	OH	H	-CHCH ₂ CO ₂ C ₂ H ₅



53 – 58

No.	R ₁	R ₂	R ₃
53	OH	H	H
54	H	H	OCH ₃
55	H	H	NO ₂
56	H	NO ₂	H
57	H	H	OH
58	H	OH	H



51

No.	
59	
60	
61	

Scheme. 1 (continued).

Compounds used in this study are structurally diverse; they were selected to cover reasonably broad activity range and to display a good distribution of activity values (Fig. 1). The biological activities were converted into the corresponding pIC_{50} values, as shown in Table 1.

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

2.2. Computational details

Due to the fact that bioactive conformations of these inhibitors are not known, the lowest energy conformations were reasonable initial structures to perform QSAR calculations.

Hence, the structures taken from conformational search were used in the present QSAR investigations.

All 3D structures were built and the conformational search was performed with the standard TRIPOS force field [32] using SYBYL 6.8 [33]. The resulting geometries were subsequently optimized by *ab initio* method at the HF/3-21G level of theory, implemented in Gaussian98 program package [34]. The 3D-QSAR calculations were then performed using SYBYL 6.8.

2.2.1. Alignment method

The bioactive conformation of **1**, (5-chloroindol-3-yl)-3-hydroxy-3-(2H-tetrazole-5-yl)-propenone, 5CITEP, was extracted

from the X-ray crystallographic data of the HIV-1 IN/5CITEP complex, available in the Protein Data Bank (1QS4) [35]. Its C1–C5 atoms (Fig. 2) were used as a template for superimposition of all molecules. Although compounds used in this study exhibit significant structural diversity, each compound has substructure (as labeled by the numbers 1–5 in Schemes 1 and 2) that corresponds to atoms C1–C5 of 5CITEP and can fit well to those atoms, as illustrated by the superimposition of all compounds (Fig. 3).

Generally, the results of CoMFA analyses depend upon the alignment method of molecules. Although the comparisons of different alignment techniques were reported [11,36] there is no preference for certain alignment methods. Therefore, three different alignment techniques were compared carefully in this work, in order to find the most efficient one for the present system.

2.2.1.1. The atom-based rms fit method. In this method, each molecule was superimposed to the template by minimizing the rms distances between each pair of corresponding atoms of the template and the compound to be aligned.

2.2.1.2. The flexible fitting (multi-fit) method. Here, alignments of molecules were performed by a multi-fit option allowing flexible fitting of the molecules to the template.

2.2.1.3. The rigid body field fit method. In this field fit procedure, the rms differences between the fitting molecule and the template molecule for steric and electrostatic fields averaged across all lattice points were minimized.

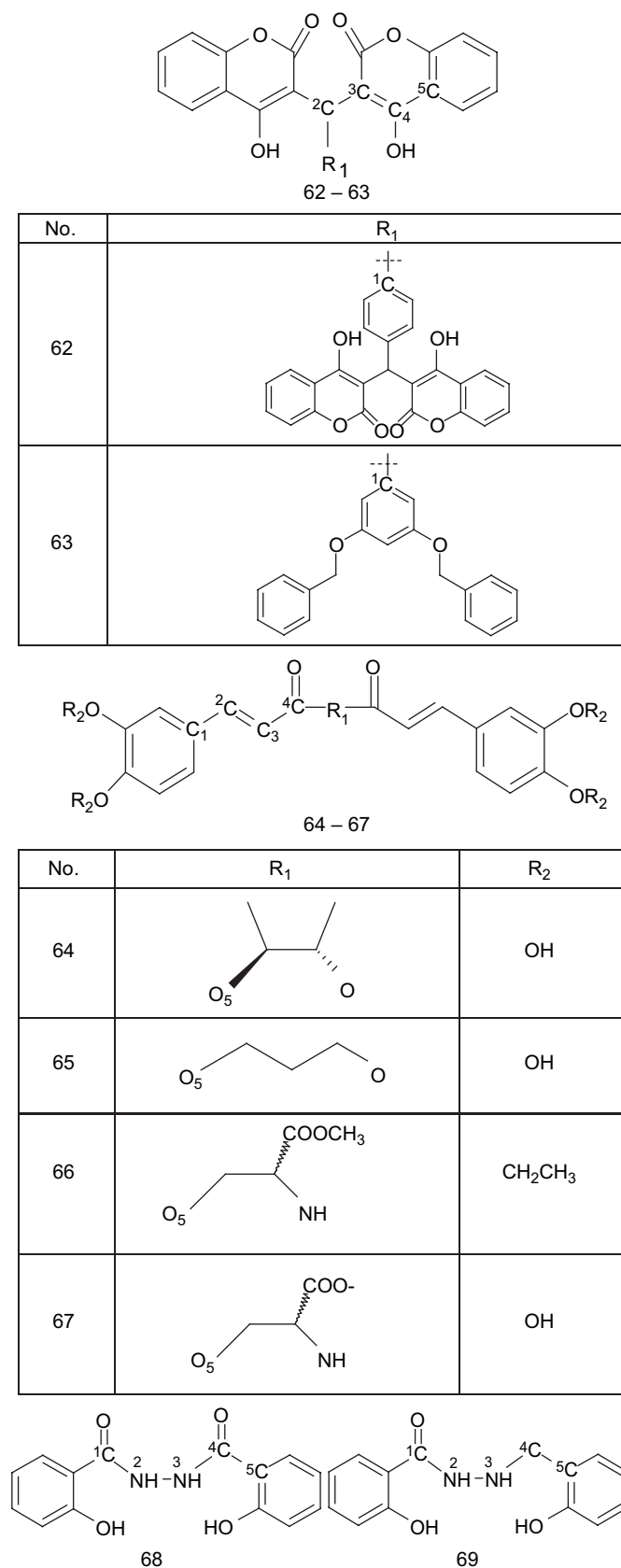
2.2.2. CoMFA calculations

CoMFA calculations were carried out using the default setup. CoMFA calculates steric fields using the Lennard–Jones potentials, and electrostatic fields using the Coulomb potentials. The CoMFA region was defined by extending the van der Waals radii of the assembly of superimposed molecules by 4 Å along the three axes of the Cartesian coordinate system. The grid spacing was set to 2.0 Å. Both the steric and the electrostatic fields were calculated for each molecule using a carbon sp^3 probe atom with a charge of +1, and energy cut-off was set to 30 kcal/mol. The partial atomic charges for each compound were assigned by the Gasteiger–Marsili method [37], implemented in the SYBYL 6.8. The analyses were performed with a scaling according to CoMFA standard deviations.

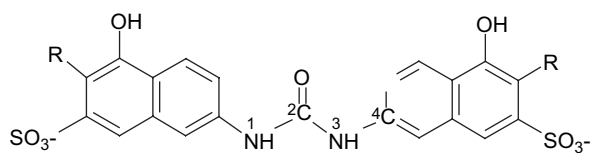
CoMFA region focusing method [38], an iterative procedure which refines a model by increasing the weights of the lattice points, was further applied in order to enhance the predictability of a CoMFA study. StDev \times coefficient values were used as weights.

2.2.3. CoMSIA calculations

In the present CoMSIA investigations, five different similarity fields including steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor interactions were calculated. The previous grid generated for the CoMFA study

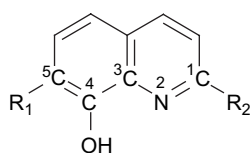


Scheme 2. Structures of compounds used in the test set. The labeled numbers correspond to atoms C1–C5 of 5CITEP (template) used in superposition process.



70 – 74

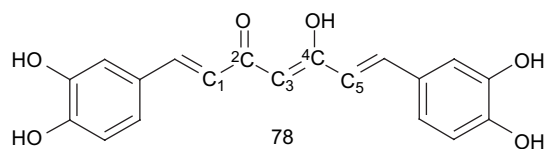
No.	R
70	
71	
72	
73	
74	



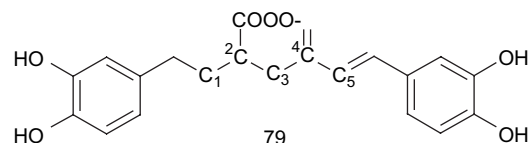
75 – 77

No.	R ₁	R ₂
75	COO ⁻	
76	COO ⁻	
77	COO ⁻	

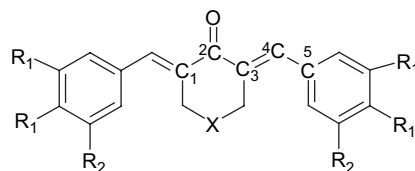
Scheme. 2 (continued).



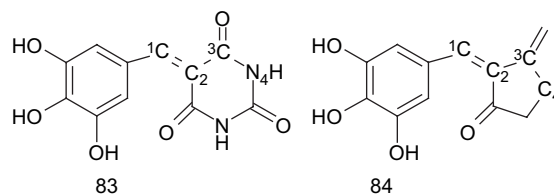
78



79



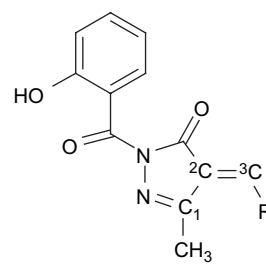
80 – 82, 85



83

84

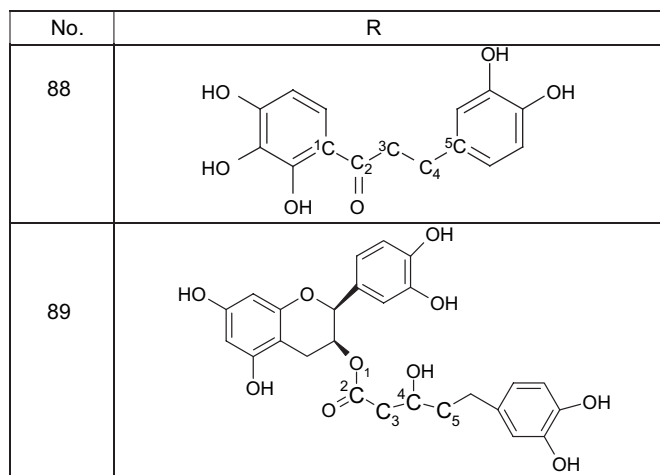
No.	R ₁	R ₂	X
80	Cl	H	O
81	OH	H	CH ₂
82	OH	H	O
85	OH	OH	CHCH ₂ CO ₂ CH ₂ CH ₃



86 – 87

No.	R
86	
87	

Scheme. 2 (continued).



Scheme. 2 (continued).

was also used for CoMSIA field calculations. An sp^3 carbon probe atom with a charge of +1, a hydrophobicity value +1, hydrogen bond donor and hydrogen bond acceptor properties set to +1 was placed at every grid point to measure the electrostatic, steric, hydrophobic, hydrogen bond donor and hydrogen bond acceptor fields, respectively. Gaussian type distance dependence and the default value of the attenuation factor ($\alpha = 0.3$) were used.

2.2.4. Partial least square analysis

After all CoMFA and CoMSIA fields were calculated, partial least square (PLS) analyses [39,40] were carried out. Cross-validations [41,42] in PLS were performed by leave-one-out (LOO) procedure. The overall quality of the analyses was expressed by the cross-validated r^2 (r_{cv}^2) values. The optimum number of components (ONC) which corresponds to the highest r_{cv}^2 and the lowest standard error of prediction (SEP) was evaluated. To speed up the analysis and to reduce the amount of noise, column filtering was set to 2.0 kcal/mol. Subsequently, PLS analyses were performed without cross-validation, using the ONC obtained from the former cross-validation procedure. Conventional correlation coefficients (r^2) and their standard errors of estimate (SEE) were computed. The

Table 1

Actual and calculated activities of compounds used in training and test sets, obtained from CoMFA and CoMSIA

No.	Actual pIC_{50}	Predicted pIC_{50}	
		CoMFA	CoMSIA
Training set			
1	4.76	4.75	4.74
2	5.16	5.24	5.15
3	5.16	5.23	5.12
4	5.00	5.16	5.01
5	4.82	4.94	4.97
6	4.70	4.53	4.71
7	4.46	4.54	4.68
8	4.37	4.74	4.70
9	4.27	4.30	4.76
10	4.62	4.68	4.92
11	5.41	5.14	4.99
12	5.59	5.53	5.28
13	5.54	5.55	5.28
14	4.98	4.77	4.70
15	5.01	5.14	5.11
16	5.00	5.15	5.20
17	5.68	5.86	5.69
18	5.17	5.17	5.19
19	5.38	5.38	5.47
20	4.96	5.09	5.05
21	5.64	5.40	5.52
22	5.04	4.93	5.33
23	5.40	5.37	5.61
24	6.46	6.36	5.77
25	6.00	5.94	5.88
26	5.70	5.61	5.77
27	5.57	5.56	5.74
28	5.40	5.21	5.75
29	6.05	5.80	5.80
30	6.05	5.72	5.88
31	6.16	6.01	6.07
32	6.59	6.21	6.07
33	5.47	5.60	5.38
34	5.39	5.56	5.48
35	5.46	5.64	5.66
36	5.80	5.66	5.56
37	5.43	5.55	5.85
38	5.62	5.68	5.99
39	5.55	5.66	5.58
40	4.74	4.72	4.58
41	4.43	4.40	4.75
42	4.70	4.52	4.69
43	4.52	4.40	4.41
44	4.52	4.57	4.64
45	5.05	5.10	4.76
46	6.30	6.21	6.29
47	6.30	6.21	6.30
48	5.77	5.70	5.56
49	5.22	5.21	5.26
50	6.16	6.27	6.27
51	5.80	5.63	5.49
52	5.55	5.41	5.68
53	6.22	6.18	5.97
54	6.05	6.17	6.03
55	6.10	6.17	6.07
56	5.85	5.97	6.06
57	6.22	6.20	6.05
58	6.05	6.26	6.10
59	4.62	4.67	4.39
60	4.55	4.61	4.45

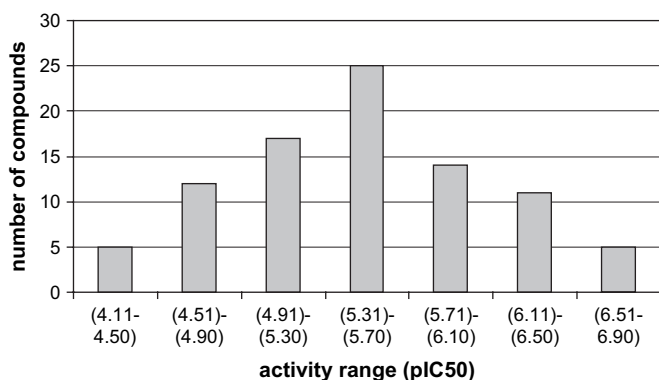


Fig. 1. Distribution plot of activity values of the training set.

Table 1 (continued)

No.	Actual <i>p</i> IC50	Predicted <i>p</i> IC50	
		CoMFA	CoMSIA
Training set			
61	4.46	4.55	4.36
Test set			
62	5.82	5.12	5.12
63	5.00	4.58	4.57
64	4.62	4.93	4.91
65	4.66	4.71	5.34
66	5.60	5.46	5.49
67	5.48	5.12	5.22
68	5.68	5.09	5.34
69	5.17	5.29	4.48
70	6.52	6.23	6.04
71	6.40	6.34	6.03
72	6.30	6.21	6.00
73	6.30	6.12	5.83
74	6.30	6.07	5.90
75	5.28	5.54	5.37
76	5.49	5.70	5.85
77	6.52	5.72	5.93
78	5.22	4.71	4.50
79	5.05	5.04	4.71
80	5.05	5.22	4.80
81	6.05	6.26	6.23
82	6.70	6.08	5.76
83	5.52	5.05	4.41
84	6.52	6.12	6.04
85	5.59	5.62	6.02
86	5.70	5.37	5.08
87	5.57	5.51	5.07
88	5.77	5.07	4.89
89	5.85	5.62	5.52

CoMFA and CoMSIA results were interpreted graphically by field contribution maps using the field type “StDev \times coeff”.

3. Results and discussion

3.1. CoMFA statistics

The statistical values of CoMFA results are summarized in Table 2. The CoMFA model using the atom-based rms fit method has $r_{cv}^2 = 0.678$ with 6 components, $r^2 = 0.957$ and $r_{pred}^2 = 0.719$. The steric and electrostatic contributions are 70.5% and 29.5%, respectively. The flexible multi-fit

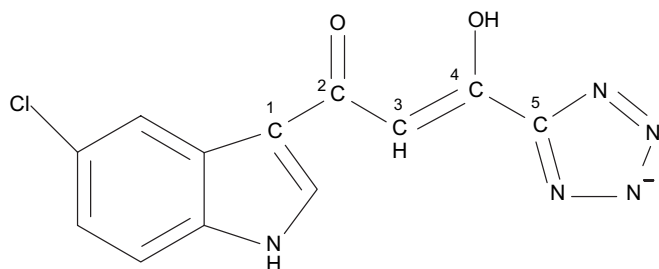


Fig. 2. Structure of 5CITEP and its atoms used for superposition are labeled.

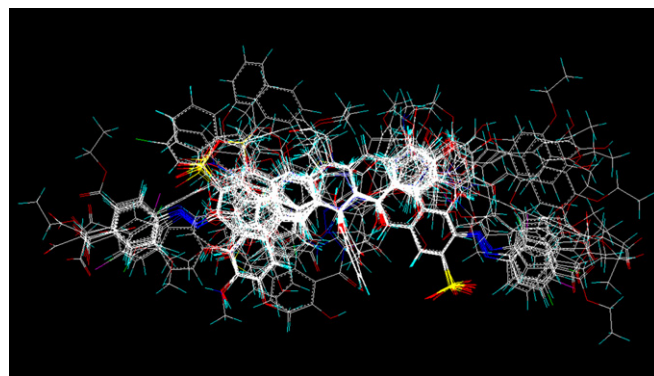


Fig. 3. 3D-view of aligned molecules (training and test sets) based on rms fit alignment method is displayed.

alignment yields $r_{cv}^2 = 0.647$ with 6 components, $r^2 = 0.939$ and $r_{pred}^2 = 0.412$. Steric and electrostatic contributions are 70.0% and 30.0%, respectively. The CoMFA model generated from field fit alignment shows a poor r_{cv}^2 of 0.546 with 6 components, $r^2 = 0.953$, and $r_{pred}^2 = 0.495$. From this alignment, the steric and electrostatic contributions are 70.6% and 29.4%, respectively. The steric and electrostatic contributions of all three models are almost similar ($\sim 70:30$), indicating more requirement of steric fields on ligand–receptor interactions. To improve the r_{cv}^2 , the region focusing technique was applied on only the atom-based rms fit model, which showed the best statistical values. After region focusing, the r_{cv}^2 of this particular CoMFA model increased from 0.678 to 0.698, whereas r^2 slightly dropped from 0.957 to 0.947. The steric and electrostatic contributions are 60.8% and 39.2%, respectively. The predictive r^2 of 0.704 was obtained. Based on r_{cv}^2 of all four models, the region focusing of the atom-based rms fit alignment with the highest r_{cv}^2 was selected to present CoMFA contour plots. The observed versus predicted activities of the training and test sets generated by a region focusing of atom-based rms fit as given in Table 1 showed good consistency.

3.2. CoMFA contours

CoMFA steric and electrostatic contours are shown in Fig. 4. In order to investigate the relationship between contours and ligand–receptor interaction, the contours were mapped onto the structure of HIV-1 IN catalytic core domain/5CITEP complex, taken from the Protein Data Bank (1QS4) [35]. The mapping was done by superimposing 5CITEP structure in the CoMFA calculations to that in the X-ray structure. The steric interaction is represented by green and yellow contours, while electrostatic interaction is denoted by red and blue contours.

Large green contour was found in a plane of the indole ring of 5CITEP indicating that bulky substituents were preferred in this region (Fig. 4a). This may be the reason why compounds with large aromatic substituents in this area, e.g. compounds 23–28 and 64–68, are more potent than molecules with small substituents, such as compounds 5–10, 40–44, 63, and 73.

Table 2
Summary of CoMFA results

Alignment	r^2_{cv}	ONC ^d	SEP ^e	r^2	SEE ^f	<i>F</i> value	r^2 pred ^g	Contributions	
								<i>S</i> ^h	<i>E</i> ⁱ
1 ^a	0.678	6	0.356	0.957	0.130	201.25	0.719	0.705	0.295
2 ^b	0.647	6	0.373	0.939	0.155	138.59	0.412	0.700	0.300
3 ^c	0.546	6	0.423	0.953	0.137	181.06	0.495	0.706	0.294
1 ^{a, j}	0.698	6	0.345	0.947	0.145	161.01	0.704	0.608	0.392

^a Alignment by rms fit.

^b Alignment by multi-fit.

^c Alignment by field fit.

^d Optimum number of components.

^e Standard error of predictions.

^f Standard error of estimates.

^g Predictive r^2 .

^h Steric.

ⁱ Electrostatic.

^j Region focusing.

This sterically preferred area is located near hydrophobic amino acids such as Phe121 and Phe139. Hence, more bulky aromatic substitutions of the inhibitors can interact better with the side chains of these residues via hydrophobic type attraction. Another steric favored region is close to the chloroindole ring of 5CITEP. The catechol moiety of the most active compound, compound **32**, is located near to this green region and therefore it exhibits higher potency than compounds without functional groups extended to this area, e.g. compound **9**. The side chains of Pro142 and Tyr143 are very close to this green contour. This is in agreement with a previous work in which residue Tyr143 was proposed to form π – π stacking or hydrophobic interactions with aromatic portion of ligands [31]. Three yellow contours are located around the tetrazole ring of 5CITEP suggesting that small bulky groups are required to increase the activity. This is possibly a reason why coumarins, compounds **1**–**14** and **62**, are less potent than styrylquinolines. The benzo-2-pyrone ring system of coumarins overlapped with one of these three yellow contours suggesting that these rings might be embedded in a narrow pocket of the enzyme, therefore, larger substituents in these regions should reduce the activity. These yellow contours are found near key residues Asn155, Lys156 and Lys159, which were reported to be involved in DNA binding process [43]. The other sterically disfavored yellow regions are located near the nitrogen atom of the indole ring of 5CITEP, close to Val113 and Thr115 residues. In addition, the sterically unfavored area was also observed above the carbonyl group of 5CITEP. This steric unfavorable contour supports again the experimental data that coumarins are less active than styrylquinolines.

The CoMFA electrostatic contour plot is displayed in Fig. 4b. Blue contours indicate that substituents should be electron deficient for high binding affinity. Since these contours were found close to indole ring of 5CITEP and to hydroxyl groups of coumarins (compounds **11**–**14**, and **62**), in which both groups are electron rich functionalities, these compounds exhibit low activities. In contrast, compound **32**, the most potent inhibitor, has no functional group with high electron density extended to these blue areas. These positively

charged favored regions were observed near Gln148 and Gly149 residues implying that electron deficient groups may interact with side chains of these residues and therefore increase the inhibitory potencies. Another blue contour lies in the region proximal to the nitrogen atom in the indole ring of 5CITEP. This contour can be used to explain the trends of the biological activities of coumarin derivatives. The NO₂ and the C₆H₅COO substituents of compounds **9** and **12**, respectively, are located in this blue region. The NO₂ substituent is a strong electron withdrawing group, whereas C₆H₅COO substituent is a weaker one. This explains why compound **9** has a lower activity than compound **12**. The Asp116 residue of the HIV-1 IN enzyme, which prefers positive charge suitable for ligand–enzyme interactions, was also observed near this blue contour.

Red contours were found near the CH adjacent to the carbonyl carbon and the hydroxyl carbon of 5CITEP, indicating a preference for negatively charged substituents in these areas. As expected, red contours are close to the Mg²⁺, which is located between Asp64 and Asp116 of HIV-1 IN. The aromatic moiety common to many IN inhibitors has been proposed to interact with this divalent cation with a cation– π type interaction [29,44], or by a charge–charge interaction between metal ion and partial or ionic charges of inhibitors [44]. Therefore, substituents with high electron density will strongly interact with Mg²⁺ leading to an enhanced biological activity of HIV-1 IN inhibitors. Additionally, the presence of negative charges favored red contours surrounding the tetrazole ring and indicates that electron rich groups may increase high binding affinity. These contours were observed near residues Asn155, Lys156 and Lys159 which favor a negative charge.

3.3. CoMSIA statistics

A total of 16 CoMSIA models were generated using either single or combined fields. The alignment giving the highest r^2_{cv} in CoMFA, atom-based rms fit method, was used. The statistical parameters are summarized in Table 3. Among the different field combinations, the CoMSIA model number 11 with

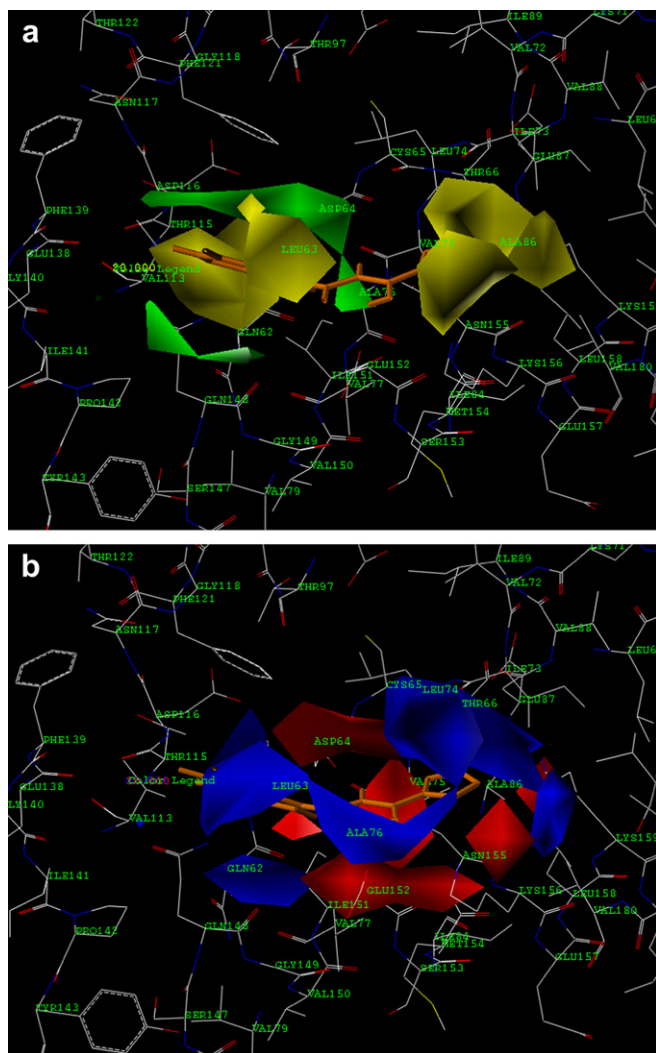


Fig. 4. CoMFA StDev \times coeff contour maps: (a) steric fields; green contours represent sterically favored regions, yellow contours indicate sterically disfavored regions, (b) electrostatic fields; red regions represent negative potential favored, blue regions indicate positive charge favored. The contours were mapped into the active site of HIV-1 IN and 5CITEP for reference. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

three fields, steric, hydrogen bond donor and hydrogen bond acceptor, gave the highest statistical values, i.e. $r_{cv}^2 = 0.693$, $ONC = 5$, $r^2 = 0.872$ and $r_{pred}^2 = 0.568$. The steric, hydrogen bond donor and hydrogen bond acceptor contributions were 24.5%, 47.4% and 26.3%, respectively, reflecting the importance of steric, the hydrogen bond donor and the hydrogen bond acceptor interactions in the inhibition mechanism. The other CoMSIA models also showed significant internal and external predictions (except the CoMSIA model number 2 which was derived from the electrostatic field only). The region focusing was applied to CoMSIA model number 11. After region focusing (model number 12), the r_{cv}^2 of this specific model increased from 0.693 to 0.724 ($ONC = 5$). The $r^2 = 0.864$ and $r_{pred}^2 = 0.524$ were obtained. The steric, hydrogen bond donor and hydrogen bond acceptor contributions are 30.3%, 43.4% and 26.3%, respectively. This CoMSIA result is in good

accordance with a recent pharmacophore study [45], which proposed that hydrogen bond donor and acceptor features were necessary descriptors for HIV-1 IN inhibitors. The actual and predicted pIC_{50} values of the training and test sets obtained from CoMSIA model number 12 are given in Table 1 and they demonstrated a good performance of our CoMSIA model.

3.4. CoMSIA contours

The CoMSIA model number 12 was used to generate contour plots. Since the steric contours of CoMSIA are very similar to those of the CoMFA (Fig. 4a), only hydrogen bonding interaction fields will be described. As the hydrogen bond fields generally contain information about the position of hydrogen bond acceptor and hydrogen bond donor groups on receptor, the contours were mapped onto the experimental IN-5CITEP complex (1QS4) [35].

Fig. 5 displays the hydrogen bond donor plot represented by cyan and purple contours. Cyan contours indicate regions where hydrogen bond donor substituents on ligands are favored and purple contours represent areas where hydrogen bond donor properties on inhibitors are disfavored. There are two cyan contours in the hydrogen bond donor maps. The first one is near the CH adjacent to the carbonyl carbon and to the hydroxyl carbon of 5CITEP, indicating that hydrogen bond donor functionalities in this region will enhance the activity. Compounds **23–28** and **64–68** are more active than 5CITEP, because they have an NH moiety located near this cyan contour. This cyan contour corresponds to Asp64 suggesting that hydrogen bond donor group of ligands may form a strong hydrogen bond with the carbonyl oxygen of this residue and hence increases inhibitory potencies. This hydrogen bond donor feature is consistent with the data obtained from a dynamic receptor based pharmacophore study [46]. The second cyan contour was observed near the indole ring of 5CITEP. Styrylquinoline analogues, compounds **29–39** and **69–71**, are more potent than 5CITEP because of the presence of hydrogen bond donating groups such as hydroxyl and methoxyl in this cyan region. Gln148 residue is located near to this cyan contour. In this case, the carbonyl group of Gln148 is hydrogen bond acceptor interacting with hydrogen bond donor substituents of ligands.

Four purple contours were found around the tetrazole ring of 5CITEP implying that the existence of hydrogen bond donor groups in this area might decrease the activity. The lower activities of coumarin analogues (compounds **1–14**, and **62**) compared to compound **32** can be further explained by the hydroxyl substituents on coumarins which lie close to these purple regions. These contours correspond to Thr66, Lys159 and Lys156 residues. As both lysines exist normally in protonated form, these amino acids are weak proton acceptors, hence, they will not interact with hydrogen bond donor groups on ligands. This finding of unfavorable region of hydrogen bond donor contour is concurrent with the previous CoMSIA study on cinnamoyls [17].

Fig. 6 shows the CoMSIA hydrogen bond acceptor field, denoted by magenta and red contours. Magenta contours

Table 3
Summary of CoMSIA results

Model	Fields	r_{cv}^2	ONC ^f	SEP ^g	r^2	SEE ⁱ	F value	r^2 pred ^j	Contributions				
									S^a	E^b	H^c	D^d	A^e
1	S	0.582	5	0.402	0.879	0.216	79.99	—	1.000	—	—	—	—
2	E	0.004	1	0.600	0.207	0.535	15.37	—	—	1.000	—	—	—
3	H	0.426	5	0.392	0.846	0.244	60.52	—	—	—	1.000	—	—
4	D	0.611	6	0.501	0.794	0.285	34.69	—	—	—	—	1.000	—
5	A	0.330	3	0.471	0.601	0.386	28.59	—	—	—	—	—	1.000
6	S + E	0.268	3	0.523	0.336	0.407	23.77	0.605	0.526	0.474	—	—	—
7	S + E + H	0.449	6	0.466	0.922	0.175	106.33	0.591	0.265	0.272	0.464	—	—
8	S + E + D	0.676	6	0.358	0.917	0.181	99.37	0.553	0.276	0.234	—	0.490	—
9	S + E + A	0.446	6	0.460	0.882	0.216	66.97	0.588	0.408	0.326	—	—	0.266
10	D + A	0.604	3	0.385	0.777	0.289	66.18	0.407	—	—	—	0.602	0.398
11	D + A + S	0.693	5	0.345	0.872	0.223	74.84	0.568	0.245	—	—	0.474	0.281
12	D + A + S ^k	0.724	5	0.327	0.864	0.231	69.59	0.524	0.303	—	—	0.434	0.263
13	D + A + E	0.578	5	0.404	0.850	0.241	62.46	0.207	—	0.225	—	0.482	0.294
14	D + A + H	0.664	6	0.364	0.923	0.174	108.62	0.454	—	—	0.353	0.356	0.291
15	S + E + D + A	0.645	5	0.371	0.896	0.201	94.35	0.574	0.182	0.183	—	0.399	0.236
16	All	0.648	6	0.373	0.936	0.158	132.51	0.566	0.143	0.154	0.234	0.283	0.186

^a Steric.

^b Electrostatic.

^c Hydrophobic.

^d Hydrogen bond donor.

^e Hydrogen bond acceptor.

^f Optimum number of components.

^g Standard error of predictions.

ⁱ Standard error of estimates.

^j Predictive r^2 .

^k Region focusing.

represent regions where hydrogen bond acceptors on ligands are favorable and red contours indicate regions where hydrogen bond acceptors on inhibitors are unfavorable for the activity.

There is one magenta contour near the keto-enol moiety of 5CITEP, which means that a hydrogen bond acceptor group in

this region will enhance inhibition potency. For instance, salicylhydrazine analogues, compounds **53–58** and **76–77**, are more potent than 5CITEP, because these compounds have carbonyl oxygens, as hydrogen bond acceptor groups, close to this magenta contour. The high activity of compounds

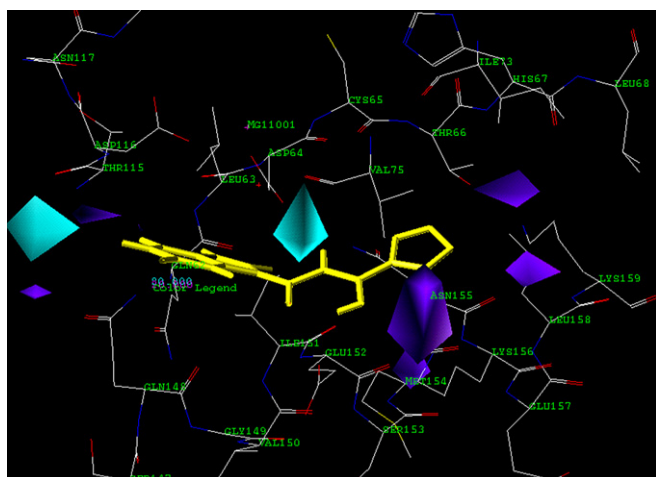


Fig. 5. Mapping of CoMSIA StDev \times coeff hydrogen bond donor contour plots within the active site of the complex structure of HIV-1 IN/5CITEP. Cyan contours represent areas where hydrogen bond donor groups on ligand enhance activity whereas purple contours indicate area where hydrogen bond donor groups on ligand decrease activity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

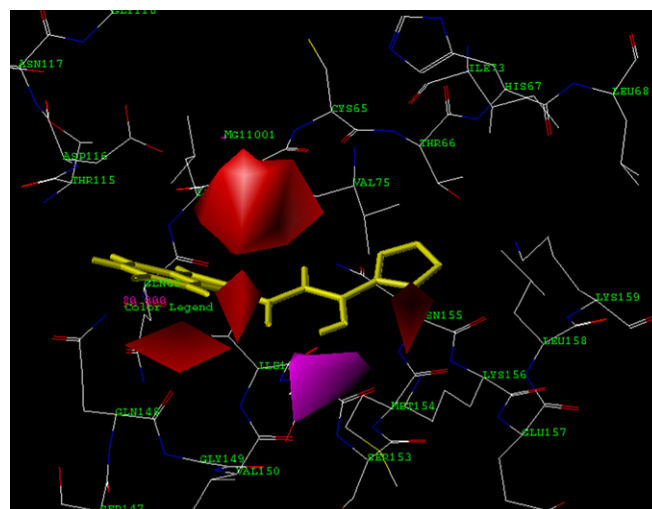


Fig. 6. Mapping of CoMSIA StDev \times coeff hydrogen bond acceptor contour plots within the active site of the complex structure of HIV-1 IN/5CITEP. Magenta contours represent regions where hydrogen bond acceptor groups on ligand enhance activity while red contours indicate areas where hydrogen bond acceptor groups on ligand decrease activity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

23–28 and **64–68** arises from the fact that oxygen atoms of SO_3^- substituents are near to this magenta area. The hydrogen bond acceptor feature on the keto-enol moiety of 5CITEP confirms the recent report on pharmacophore model of diketo acid analogues [47]. The Glu152 residue was found close to this magenta contour implying that hydrogen bond forming between ligands and this amino acid will increase binding affinities. A large red contour was found behind the NH moiety of the indole ring of 5CITEP suggesting that hydrogen bond acceptor features of ligands in this area may reduce the activity. This contour was located near Asp64 which contains $\text{C}=\text{O}$ (hydrogen bond acceptor group), indicating unfavorable hydrogen bonding affinity. The contour that disfavors donors on the receptor site is attributed to low affinity of compounds that have hydrogen bond acceptor groups in the vicinity.

4. Conclusion

In this study, we successfully derived one CoMFA model and one CoMSIA model for all 89 compounds with high structural diversity. Both models show good correlation between predicted and observed inhibitory potencies against HIV-1 IN, as indicated by r_{cv}^2 of 0.698 for CoMFA and of 0.724 for CoMSIA. CoMFA investigations were performed using three different fitting methods for alignment process. There is no significant difference between the CoMFA models derived by rms fit and multi-fit alignment methods, whereas the model derived from field fit method gave a poor r_{cv}^2 (0.546). The CoMFA results suggest that steric interactions (60.8%) as well as electrostatic interactions (39.2%) contributed to the activities of inhibitors. From the obtained results, it can be concluded that larger substituents in a plane of the indole ring and small bulky groups at the tetrazole ring of 5CITEP are required to increase the inhibitory potency. Moreover, groups of higher electron density of the ligand are necessary for a better interaction with the metal ion in the active site of enzyme.

The hydrogen bond donor and the hydrogen bond acceptor fields obtained by CoMSIA show the significance of hydrogen bond interactions between ligands and HIV-1 IN enzyme. The hydrogen bond donor and acceptor fields can be mapped back to the structure of the enzyme and they are consistent with the experimentally observed hydrogen bond between Asn155, Lys156 and Lys159 with the side chains of inhibitors. This provides some understanding on the influence of hydrogen bond interactions between enzyme and inhibitor molecules. The information obtained from CoMFA and CoMSIA could lead to a better design of suitable selective and higher potent HIV-1 IN inhibitors.

Acknowledgements

Financial support from the Royal Golden Jubilee Ph.D. Program (3.C.CU/45/S.1), the Thailand Research Fund and a generous supply of computational time and SYBYL 6.8 by the ZID of the University of Vienna, Vienna, Austria, are gratefully acknowledged. We also thank the Austrian Thai Center for Computer Assisted Chemical Education and Research (ATC),

Department of Chemistry, Faculty of Science, Chulalongkorn University, and the National Electronics and Computer Technology (NECTEC) for computer and SYBYL facilities.

References

- [1] G. Barbaro, A. Scozzafava, A. Mastrolorenzo, C.T. Supuran, *Curr. Pharm. Des.* 11 (2005) 1805–1843.
- [2] Y. Yazdanpanah, D. Sissoko, M. Egger, Y. Mouton, M. Zwahlen, G. Chene, *BMJ* 328 (2004) 249–253.
- [3] A.A. Johnson, C. Marchand, Y. Pommier, *Curr. Top. Med. Chem.* 4 (2004) 1059–1077.
- [4] F.D. Bushman, R. Craigie, *Proc. Natl. Acad. Sci. U. S. A.* 88 (1991) 1339–1343.
- [5] R. Zheng, T.M. Jenkins, R. Craigie, *Proc. Natl. Acad. Sci. U. S. A.* 93 (1996) 13659–13664.
- [6] A. Engelman, R. Craigie, *J. Virol.* 66 (1992) 6361–6369.
- [7] A. Engelman, A.B. Hickman, R. Craigie, *J. Virol.* 68 (1994) 5911–5917.
- [8] R.D. Cramer III, D.E. Patterson, J.D. Bunce, *J. Am. Chem. Soc.* 110 (1988) 5959–5967.
- [9] G. Klebe, U. Abraham, T. Mietzner, *J. Med. Chem.* 37 (1994) 4130–4146.
- [10] M.T. Makhija, V.M. Kulkarni, *J. Comput. Aided Mol. Des.* 15 (2001) 961–978.
- [11] M.T. Makhija, V.M. Kulkarni, *J. Comput. Aided Mol. Des.* 16 (2002) 181–200.
- [12] 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.
- [13] H. Yuan, A.L. Parrill, *J. Mol. Graph. Model.* 23 (2004) 317–328.
- [14] H. Yuan, A.L. Parrill, *Bioorg. Med. Chem.* 10 (2002) 4169–4183.
- [15] M.T. Makhija, V.M. Kulkarni, *Bioorg. Med. Chem.* 10 (2002) 1483–1497.
- [16] H. Yuan, A.L. Parrill, *J. Mol. Struct. (Theochem)* 529 (2000) 273–282.
- [17] J.K. Buolamwini, H. Assefa, *J. Med. Chem.* 45 (2002) 841–852.
- [18] F.F.D. Daeyaert, H.M. Vinkers, M.R. de Jonge, J. Heeres, L.M.H. Koymans, P.J. Lewi, P.A.J. Janssen, *Internet electron, J. Mol. Des.* 3 (2004) 528–543.
- [19] H. Zhao, N. Neamati, H. Hong, A. Mazumder, S. Wang, S. Sunder, G.W.A. Milne, Y. Pommier, T.R. Burke Jr., *J. Med. Chem.* 40 (1997) 242–249.
- [20] P.C.M. Mao, J.F. Mouscadet, H. Leh, C. Auclair, L.Y. Hsu, *Chem. Pharm. Bull.* 50 (2002) 1634–1637.
- [21] Z. Lin, N. Neamati, H. Zhao, Y. Kiryu, J.A. Turpin, C. Aberham, K. Strebel, K. Kohn, M. Witvrouw, C. Pannecouque, Z. Debyser, E. De Clercq, W.G. Rice, Y. Pommier, T.R. Burke Jr., *J. Med. Chem.* 42 (1999) 1401–1414.
- [22] H. Zhao, N. Neamati, S. Sunder, H. Hong, S. Wang, G.W.A. Milne, Y. Pommier, T.R. Burke Jr., *J. Med. Chem.* 40 (1997) 937–941.
- [23] K. Maurer, A.H. Tang, G.L. Kenyon, A.D. Leavitt, *Bioorg. Chem.* 28 (2000) 140–155.
- [24] F. Zouhiri, J.F. Mouscadet, K. Mekouar, D. Desmaele, D. Savoure, H. Leh, F. Subra, M.L. Bret, C. Auclair, J. d'Angelo, *J. Med. Chem.* 43 (2000) 1533–1540.
- [25] C. Benard, F. Zouhiri, M.N. Bayle, M. Danet, D. Desmaele, H. Leh, J.F. Mouscadet, G. Mbemba, C.M. Thomas, S. Bonnenfant, M.L. Bret, J. d'Angelo, *Bioorg. Med. Chem. Lett.* 14 (2004) 2473–2476.
- [26] A. Mazumder, N. Neamati, S. Sunder, J. Schulz, H. Pertz, E. Eich, Y. Pommier, *J. Med. Chem.* 40 (1997) 3057–3063.
- [27] M. Artico, R.D. Santo, R. Costi, E. Novellino, G. Greco, S. Massa, E. Tramontano, M.E. Marongiu, A.D. Montis, P.L. Colla, *J. Med. Chem.* 41 (1998) 3948–3960.
- [28] R. Costi, R.D. Santo, M. Artico, S. Massa, R. Ragno, R. Loddo, M.L. Colla, E. Tramontano, P.L. Colla, A. Pani, *Bioorg. Med. Chem.* 12 (2004) 199–215.
- [29] N. Neamati, H. Hong, J.M. Owen, S. Sunder, H.E. Winslow, J.L. Christensen, H. Zhao, T.R. Burke Jr., G.W.A. Milne, Y. Pommier, *J. Med. Chem.* 41 (1998) 3202–3209.

- [30] N. Neamati, H. Hong, S. Sunder, G.W.A. Milne, Y. Pommier, *Mol. Pharmacol.* 52 (1997) 1041–1055.
- [31] G.C.G. Pais, X. Zhang, C. Marchand, N. Neamati, K. Cowansage, E.S. Svarovskaia, V.K. Pathak, Y. Tang, M. Nicklaus, Y. Pommier, T.R. Burke Jr., *J. Med. Chem.* 45 (2002) 3184–3194.
- [32] M. Clark, R.D. Crammer III, N.V. Opdenbosch, *J. Comput. Chem.* 10 (1989) 982–1012.
- [33] SYBYL Molecular Modeling Softwares, Version 6.5, Tripos Associates, Inc., St. Louis, MO, 1998.
- [34] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery, Jr., R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, P. Salvador, J.J. Dannenberg, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, A.G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, J.L. Andres, C. Gonzalez, M. Head-Gordon, E.S. Replogle, J.A. Pople, Gaussian, Inc., Pittsburgh PA, 2001.
- [35] Y. Goldur, 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.
- [36] B.A. Bhongade, A.K. Gadad, *Bioorg. Med. Chem.* 12 (2004) 2797–2805.
- [37] J. Gasteiger, M. Marsili, *Tetrahedron* 36 (1980) 3219–3228.
- [38] S.J. Cho, A. Tropsha, *J. Med. Chem.* 38 (1995) 1060–1066.
- [39] W.J. Dunn III, S. Wold, U. Edlund, S. Hellberg, J. Gasteiger, *Quant. Struct.-Act. Relat.* 3 (1984) 131–137.
- [40] R.D. Cramer III, *Perspect. Drug. Discov.* 1 (1993) 269–278.
- [41] T.A. Halgren, *J. Am. Chem. Soc.* 112 (1990) 4710–4723.
- [42] B.L. Podlogar, D.M. Ferguson, *Drug. Des. Discov.* 17 (2000) 4–12.
- [43] T.M. Jenkins, D. Esposito, A. Engelman, R. Craigie, *EMBO J.* 16 (1997) 6849–6859.
- [44] M.C. Nicklaus, N. Neamati, H. Hong, A. Mazumder, S. Sunder, J. Chen, G.W.A. Milne, Y. Pommier, *J. Med. Chem.* 40 (1997) 920–929.
- [45] G.I. Mustata, A. Brigo, J.M. Briggs, *Bioorg. Med. Chem. Lett.* 14 (2004) 1447–1454.
- [46] J. Deng, K.W. Loo, T. Sanchez, M. Cui, N. Neamati, J.M. Briggs, *J. Med. Chem.* 48 (2005) 1496–1505.
- [47] M.L. Barreca, A. Rao, L.D. Luca, M. Zappala, C. Gurnari, P. Monfote, E. De Clercq, B.V. Maele, Z. Debyser, M. Witvrouw, J.M. Briggs, A. Chimirri, *J. Chem. Inf. Comput. Sci.* 44 (2004) 1450–1455.