CHAPTER IV
COMPARATIVE MOLECULAR FIELD ANALYSIS (CoMFA) AND COMPARATIVE MOLECULAR SIMILARITY INDICES ANALYSIS (CoMSIA)

3D-QSAR techniques, such as CoMFA and CoMSIA are valuable tools for discovering new drugs. As aforementioned in CHAPTER II, there were QSAR studies of several classes of HIV-1 IN inhibitors [59-64, 74, 100, 101], e.g. coumarins, quinones, hydrazides, dioxepinones, salicylpyrazolinones, compounds with catechol and without catechol moiety and mercaptobenzenesulfonamides. For calculations that included either one or a few classes of HIV-1 IN inhibitors [74, 100, 101], a single 3D-QSAR model could be achieved. However, in case of various classes of IN inhibitors [59, 61, 63], a single QSAR model describing the activities of all structurally diverse inhibitors could not be generated and all the compounds must be clustered into two groups to obtain the acceptable models. In that studies, we have to find the suitable models or equations to predict the activities of the given compounds and this is inconvenience. In this work, an attempt was made to construct a single 3D-QSAR model for variety of structural classes of HIV-1 IN inhibitors without clustering of compounds. Therefore, biological activities of any compound would be predicted by using this single model.

4.1 Computational Methods

4.1.1 Biological data

An important prerequisite for the quantitative investigation of structure activity relationships (SARs) is a suitable data set. To reduce systematic errors, bioactivity data should be ideally determined in one laboratory and uses the same method. Nevertheless, for the present work a chemically heterogeneous data set consisting of 11 and 10 classes of compounds against 3'-processing and ST process, respectively, were taken from literatures [100, 102-116]. Although the biological activities of compounds were taken from different sources, most of them (> 70%) were evaluated from the same laboratory. Likewise the previous QSAR studies [59, 61], all compounds were combined together based on the assumption that methods and conditions of activity testing were similar. Moreover, compounds used in this study are structurally diverse and they were selected

to cover reasonably broad activity range and to display a good distribution of activity values (Figure 4.1). The data analysis works better when the biological data show good distribution i.e. fairly symmetrically distributed around their mean. An unbalanced and unrepresentative set of activities yield a poor quality of 3D-QSAR and the results are difficult to interpret. The biological activities were converted into the corresponding pIC_{50} values.

$$pIC_{50} = -log\ IC_{50} \qquad (4.1)$$

For 3'-processing mechanism, the training and test sets contained 61 and 28 compounds, respectively. For ST mechanism, the training set consisted of 64 compounds and the test set comprised of 20 compounds. The structures of compounds against 3'-processing and ST mechanisms of HIV-1 IN are listed in Schemes 4.1 and 4.2, respectively. The biological activities of all compounds are given in Tables 4.2 and 4.5 for 3'-processing and ST activities, respectively.

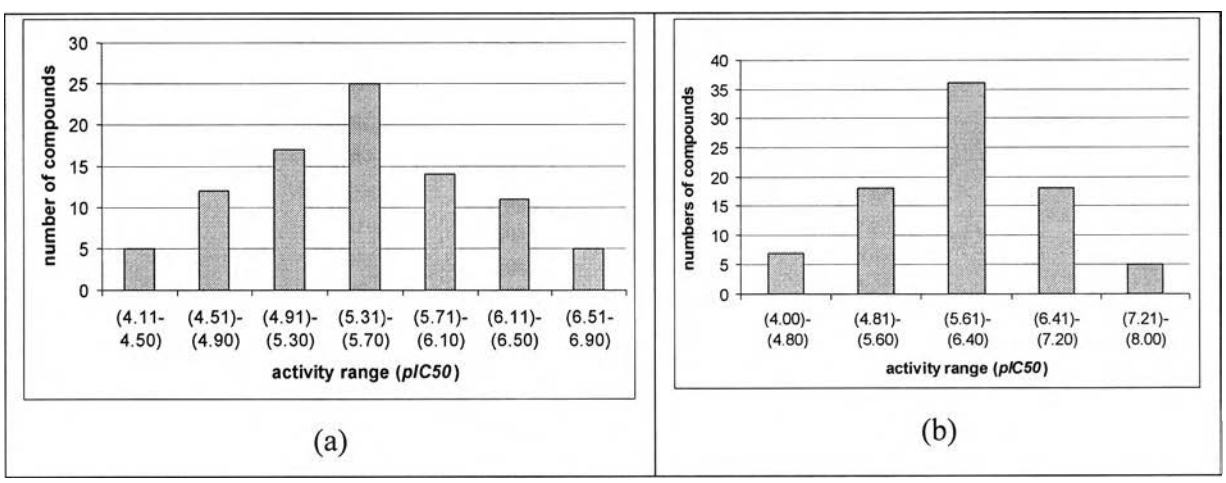
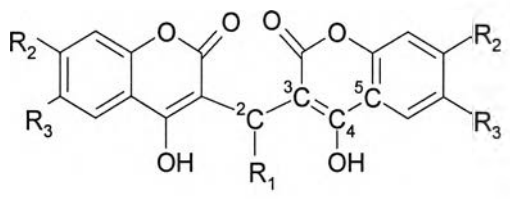
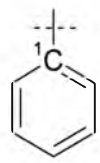
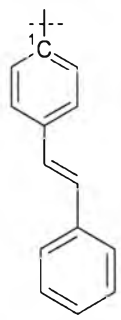
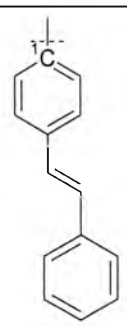


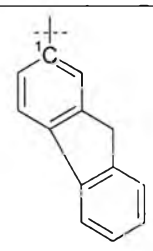
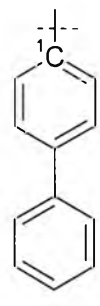
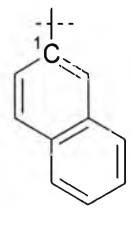
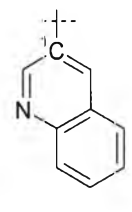
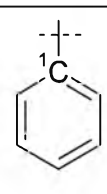
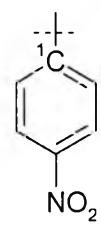
Figure 4.1 Distribution plots of activity values of the training set compounds against (a) 3'-processing and (b) ST mechanism.

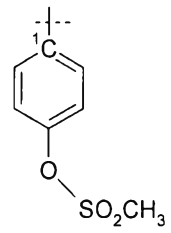
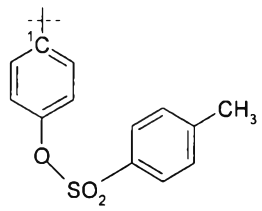
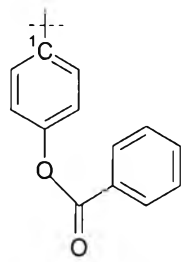
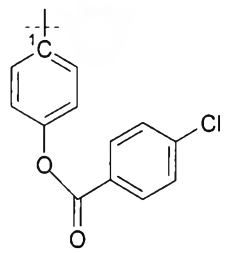
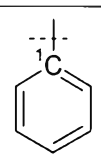
Scheme 4.1 Structures of compounds used in the training set (**1-61**) and test set (**62-89**) of 3'-processing process. The labeled numbers correspond to atoms C1-C5 of 5CITEP (template) used in superposition process.

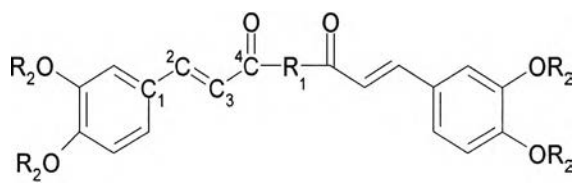


1-14 (coumarins)

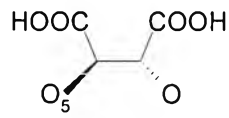
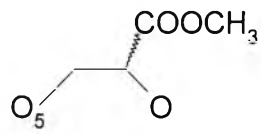
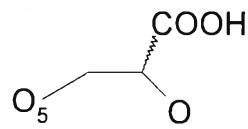
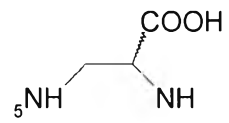
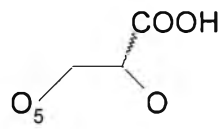
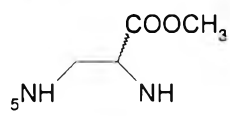
No.	R_1	R_2	R_3
1		OH	H
2		OH	H
3		H	H

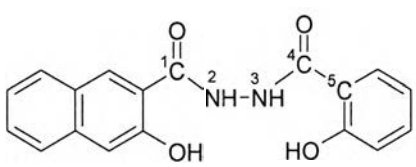
4		H	H
5		H	H
6		H	H
7		H	H
8		H	H
9		H	H

10		H	H
11		H	H
12		H	H
13		H	H
14		OH	$(CH_2)_5CH_3$

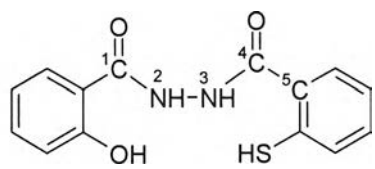


15-20 (chicoric acids)

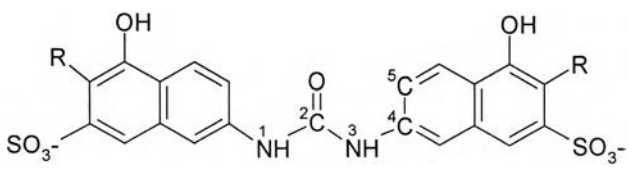
No.	R_1	R_2
15		CH_2CH_3
16		CH_2CH_3
17		CH_2CH_3
18		CH_2CH_3
19		H
20		CH_2CH_3



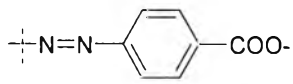
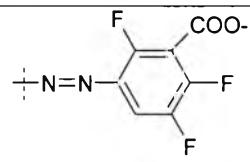
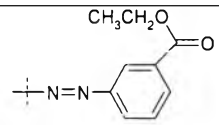
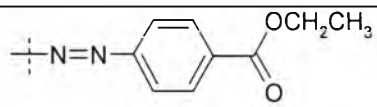
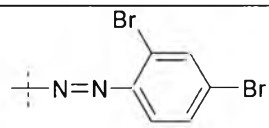
21(hydrazide)

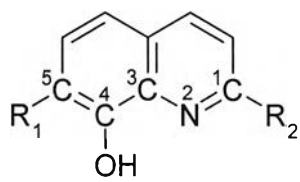


22 (hydrazide)

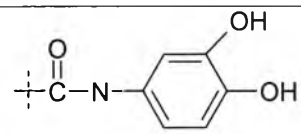
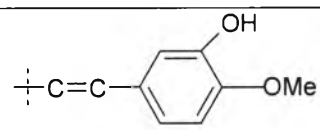


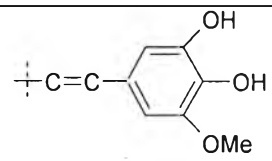
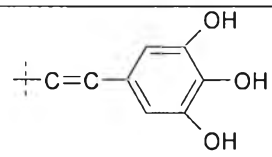
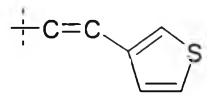
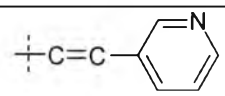
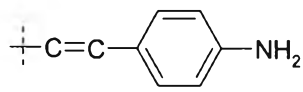
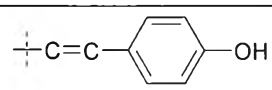
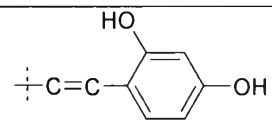
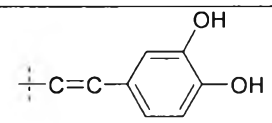
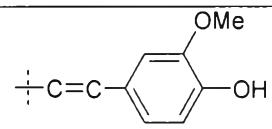
23-28 (carbonyl J derivatives)

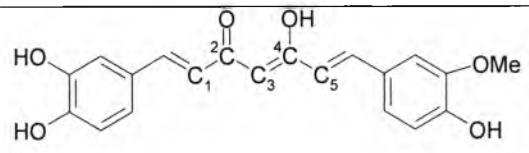
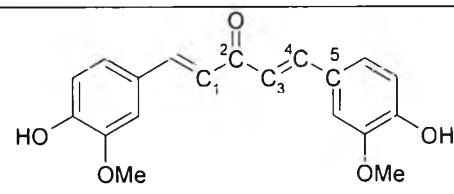
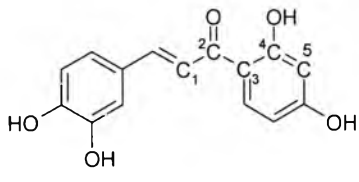
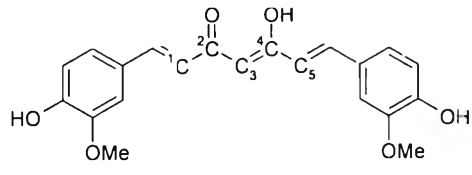
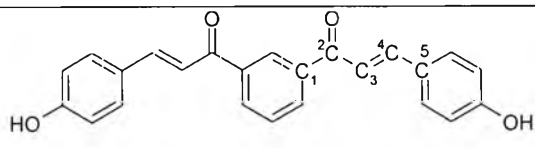
No.	R
23	H
24	
25	
26	
27	
28	

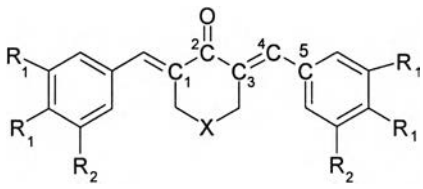


29-39 (styrylquinolines)

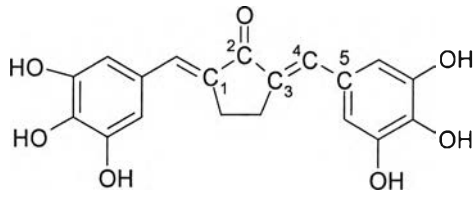
No.	R_1	R_2
29	COO^-	
30	COO^-	

31	COO^-	
32	COONa	
33	COO	
34	COO	
35	COO^-	
36	COO^-	
37	COO^-	
38	COO^-	
39	COO^-	

No.		No.	
40	 curcumin	41	 cinnamoyl
42	 cinnamoyl	43	 cinnamoyl
44	 cinnamoyl		

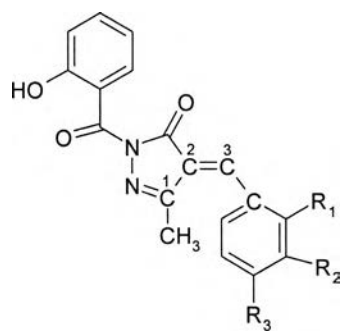


45-50, 52 (cyclohexanones)



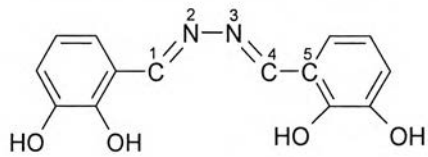
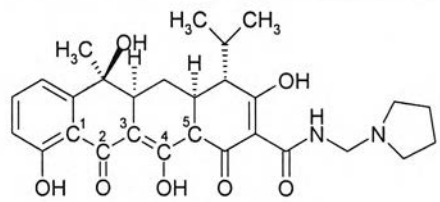
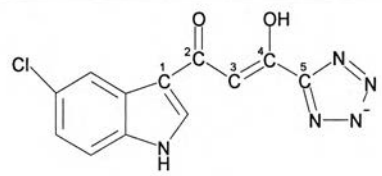
51

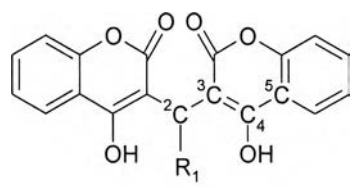
No.	R_1	R_2	X
45	Cl	H	O
46	OH	H	NCH_3
47	OH	H	NCH_2CH_3
48	OH	H	NCH_2Ph
49	OH	OH	O
50	OH	OH	S
52	OH	H	$-CHCH_2CO_2C_2H_5$



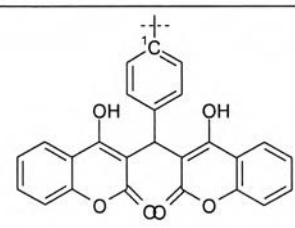
53-58 (salicylhydrazine)

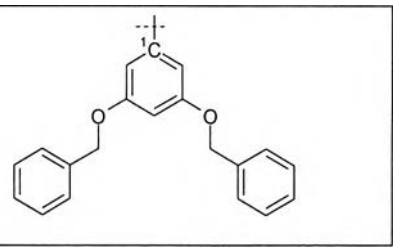
No.	R_1	R_2	R_3
53	OH	H	H
54	H	H	OCH_3
55	H	H	NO_2
56	H	NO_2	H
57	H	H	OH
58	H	OH	H

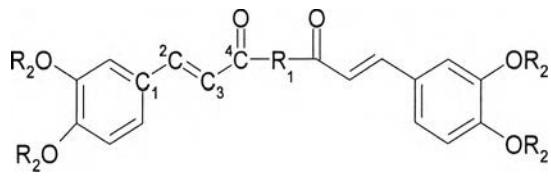
No.	
59	 tetracycline
60	 tetracycline
61	 5CITEP



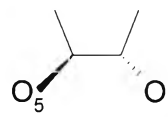
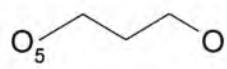
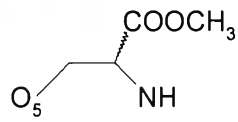
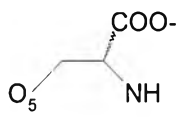
62-63

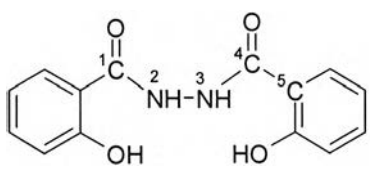
No.	R_1
62	

63	

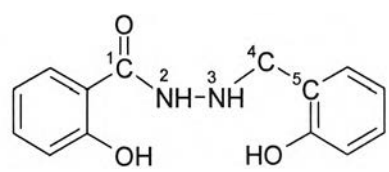


64-67

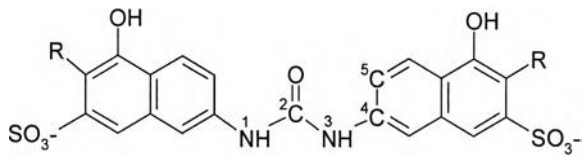
No.	R_1	R_2
64		OH
65		OH
66		CH_2CH_3
67		OH



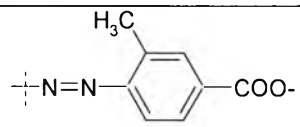
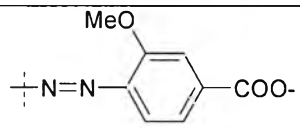
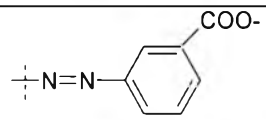
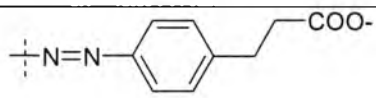
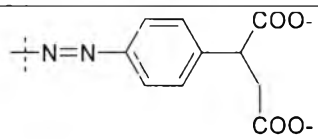
68

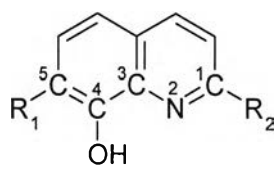


69

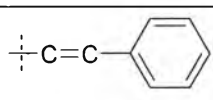
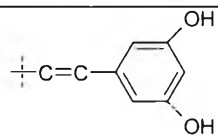


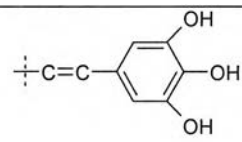
70-74

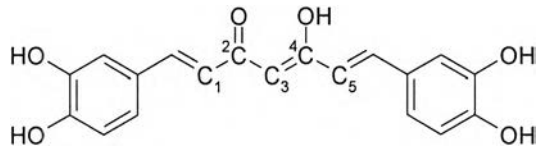
No.	R
70	
71	
72	
73	
74	



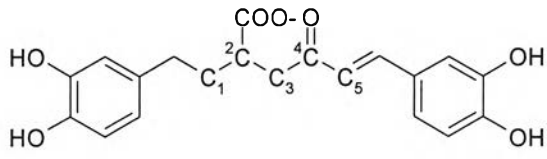
75-77

No.	R_1	R_2
75	COO^-	
76	COO^-	

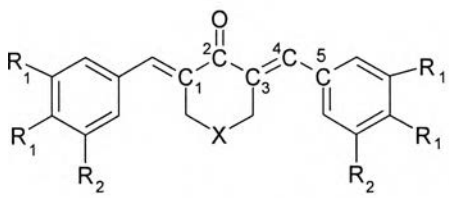
77	COO^-	



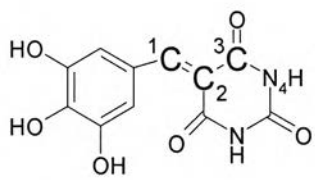
78



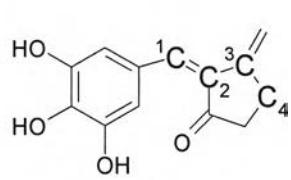
79



80-82, 85

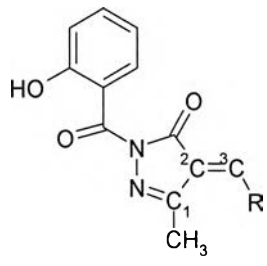


83

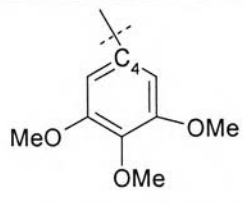
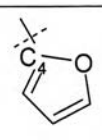
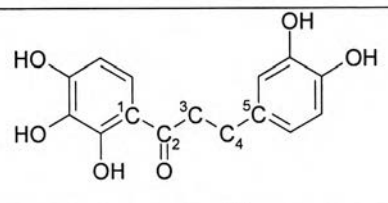
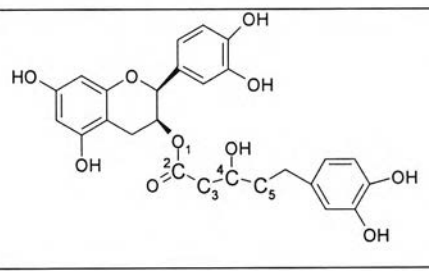


84

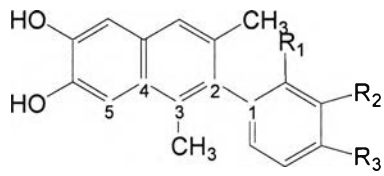
No.	R_1	R_2	X
80	Cl	H	O
81	OH	H	CH_2
82	OH	H	O
85	OH	OH	$CHCH_2CO_2CH_2CH_3$



86-87

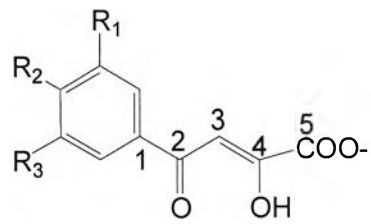
86	
87	
88	
89	

Scheme 4.2 Structures of compounds used in the training set (**1-64**) and test set (**65-84**) of ST process. The labeled numbers correspond to atoms C1-C5 of 5CITEP (template) used in superposition process.

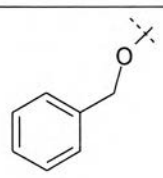
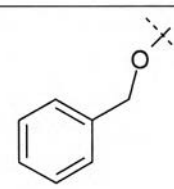


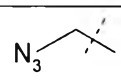
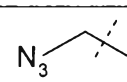
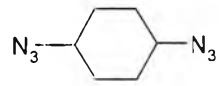
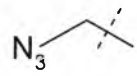
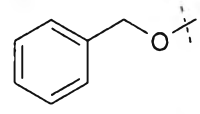
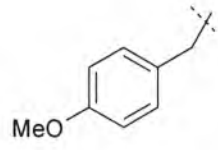
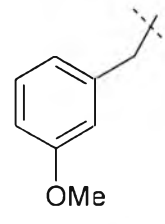
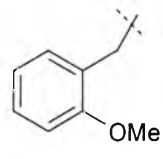
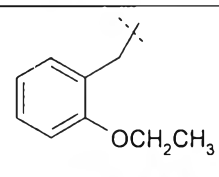
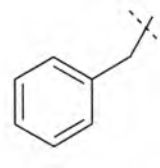
1-4 (catechol derivatives)

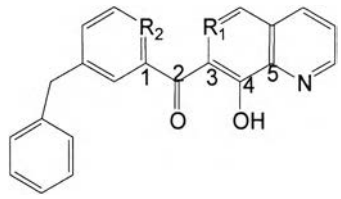
No.	R_1	R_2	R_3
1	H	CF_3	H
2	H	H	H
3	CF_3	H	H
4	H	H	CF_3



5-14 (diketo acids)

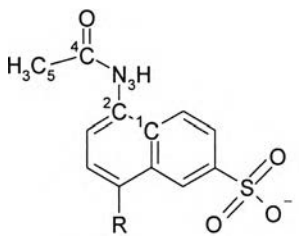
No.	R_1	R_2	R_3
5		H	

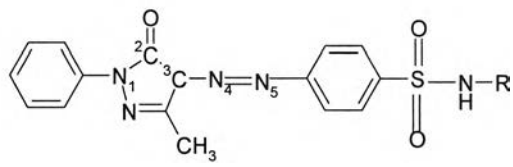
6	H		
7	H		
8	H	H	
9	H	H	
10	H	H	
11	H	H	
12	H	H	
13	H	H	
14	H	H	



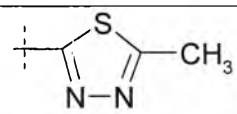
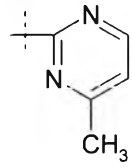
15-17 (naphthyridines)

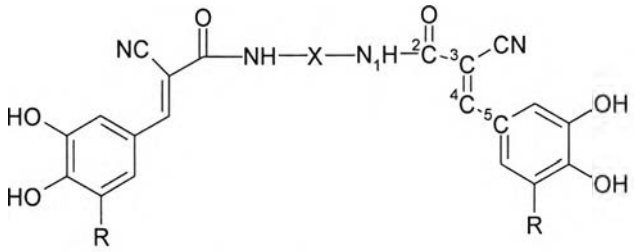
No.	R_1	R_2
15	N	C
16	C	C
17	C	N



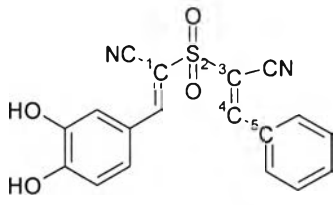


18-19 (phenylsulfones) 20-21(phenylsulfones)

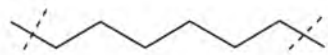
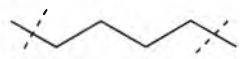
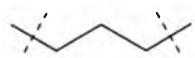
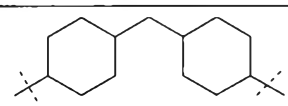
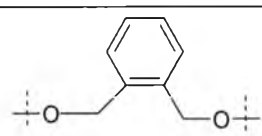
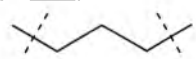
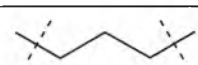
No.	R
18	H
19	NH2
20	
21	

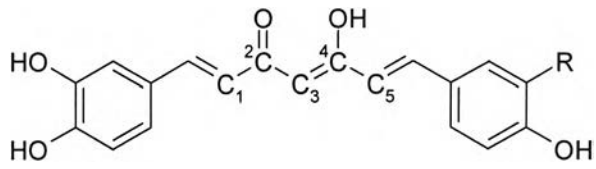


22-29 (tyrphostins)



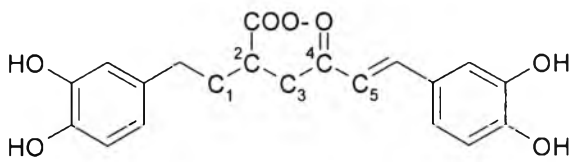
30 (tyrphostins)

No.	R	X
22	H	
23	OH	
24	Br	
25	H	
26	H	
27	NO_2	
28	OH	
29	OH	

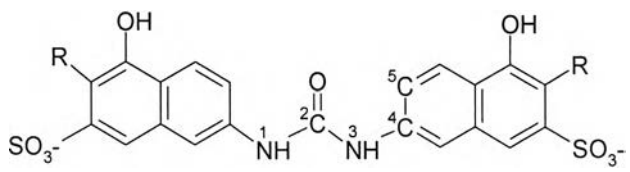


31, R= OH, (curcumin)

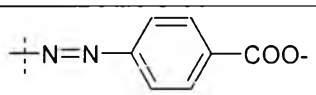
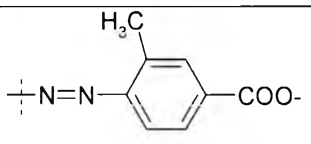
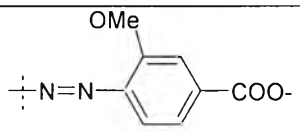
32, R= OCH_3 (curcumin)

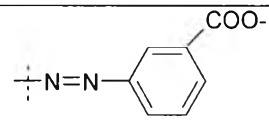
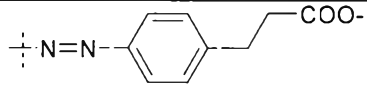
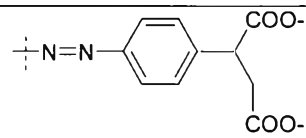
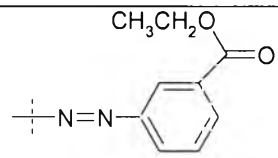
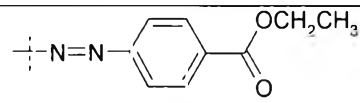
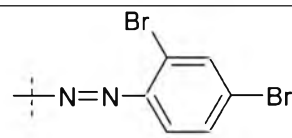


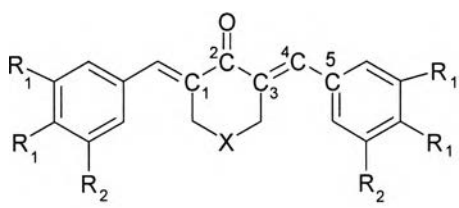
33 (curcumin)

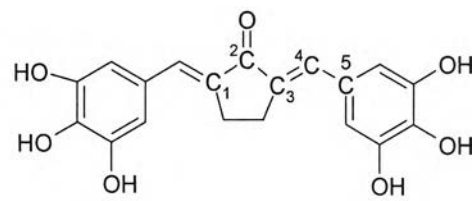


34-43 (carbonyl J derivative)

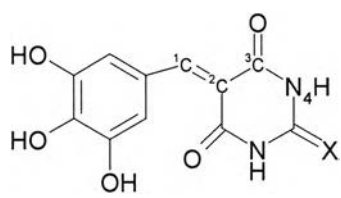
No.	R
34	H
35	
36	
37	

38	
39	
40	
41	
42	
43	



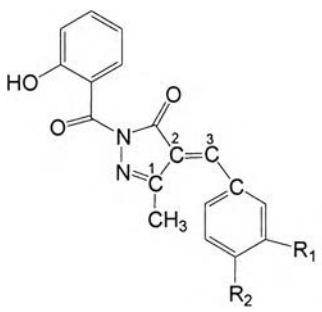


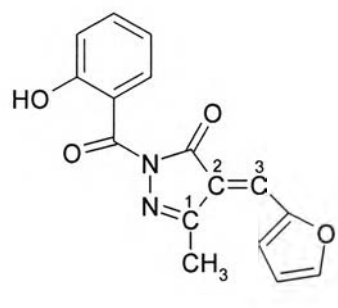
44-56 (cyclohexanones) 51



52-53

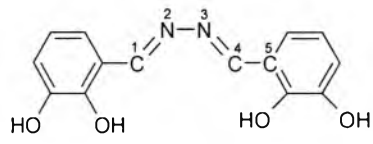
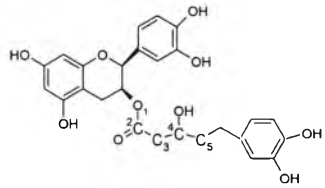
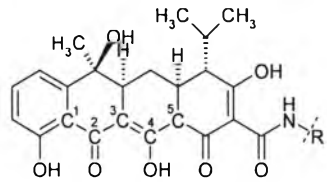
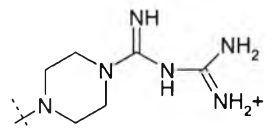
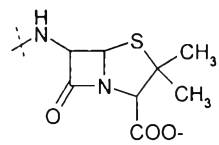
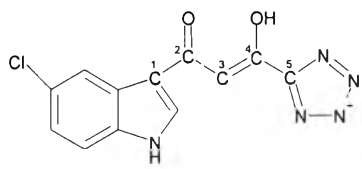
No.	R_1	R_2	X
44	Cl	H	O
45	OH	H	NCH_3
46	OH	H	NCH_2CH_3
47	OH	H	NCH_2Ph
48	OH	OH	CH2
49	OH	OH	O
50	OH	OH	S
52	-	-	O
53	-	-	S
54	OH	OH	$-COOCH_2CH_3$
55	OH	OH	-COO-
56	OH	OH	$-CH_2COO-$

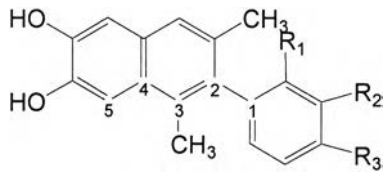




57-58 (salicylhydrazines) 59

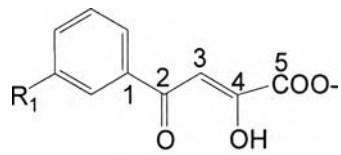
No.	R_1	R_2
57	H	NO_2
58	OCH_3	OH

No.	
60	 (tetracyclines)
61	 (tetracyclines)
	 62-63 (tetracyclines)
62	R= 
63	R= 
64	 (5CITEP)

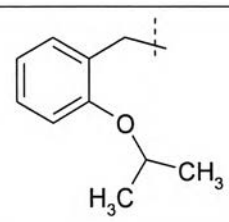
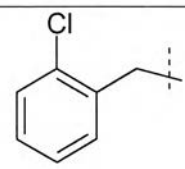


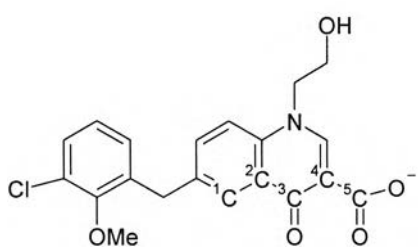
65-66

No.	R_1	R_2	R_3
65	CF_3	H	H
66	H	OH	OH

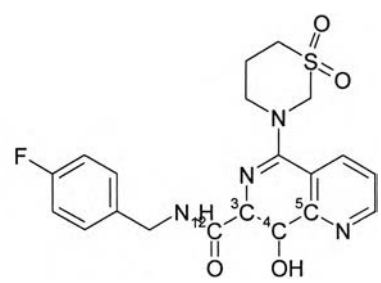


67-68

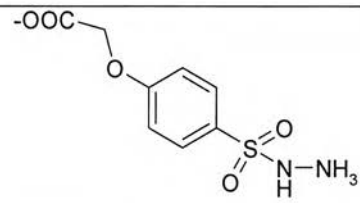
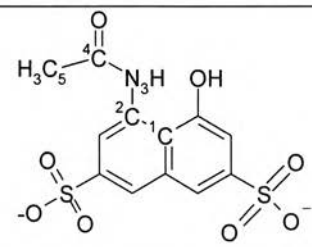
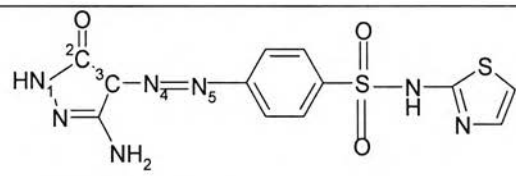
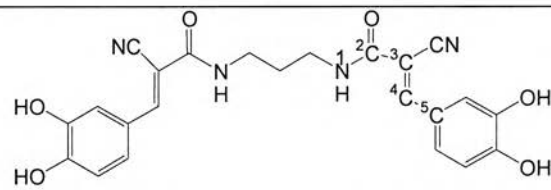
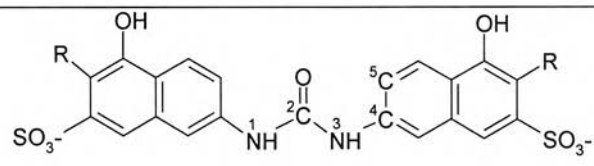
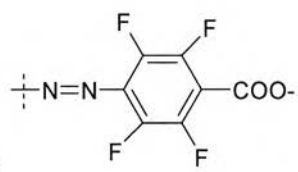
No.	R_1
67	
68	

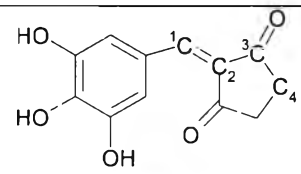
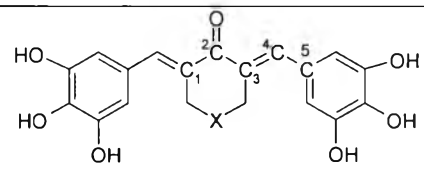


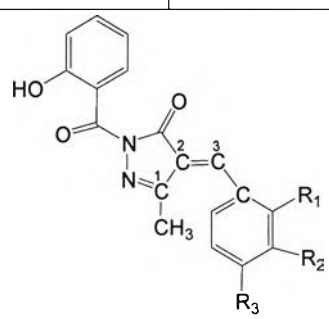
69 (JTK-303)



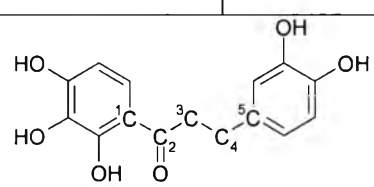
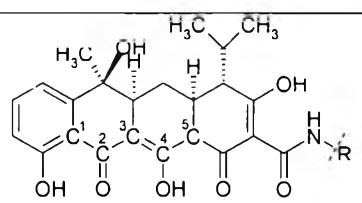
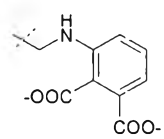
70 (L870, 810)

No.	
71	
72	
73	
74	
75	  R=

No.	
76	
77	 X= $CHCH_2CO_2CH_2CH_3$



78-82

No.	R_1	R_2	R_3
78	OH	H	H
79	H	NO_2	H
80	H	H	OH
81	H	OH	H
82	H	OCH_3	OCH_3
83			
84	 R= 		

4.1.2 Compound Generation

Ideally, biologically active forms of compounds should be used to derive 3D-QSAR models. However, because of the fact that bioactive conformations of the studied molecules are not known, the lowest energy conformations are reasonable initial structure to perform 3D-QSAR calculations.

Compounds under investigation were built and the conformational search was performed with the standard TRIPOS force field [117] implemented in SYBYL 6.8 [118]. Using the systematic search protocol, rotatable chemical bonds in compounds were searched from 0 to 360 degrees in 10 degrees increments. The lowest energy conformation found for each structure was subsequently geometry optimized using the HF/3-21G level of theory implemented in Gaussian03 [119] and consequently used in the 3D-QSAR calculations.

4.1.3 Template selection

In the derivation of 3D-QSAR models, the choice of the template conformation is important to provide the illustration of a reliable model. In principle, the bioactive conformation (exiting in the protein-ligand complex) of ligands is normally used as a template. Therefore, in this study, 5CITEP which was extracted from the X-ray complex structure of HIV-1 IN-5CITEP (pdb entry code: 1QS4 [39]) was selected as the template.

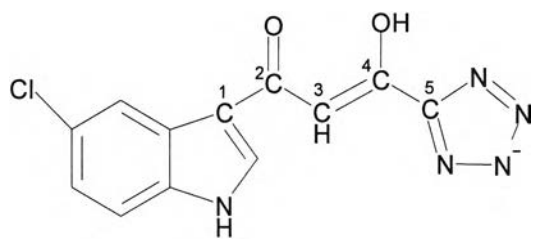


Figure 4.2 Structure of 5CITEP showing the C1-C5 atoms used for superimposition.

4.1.4 Alignment Procedure

Although compounds used in this study exhibit significant structural diversity, each of them shares similar substructure (as labeled by the numbers 1 - 5 in Scheme 1

and 2) corresponding to atom C1-C5 of 5CITEP (Figure 4.2). Therefore, these defined atoms of each molecule were fitted to those corresponding C1 - C5 of 5CITEP.

The superimposition of molecules is one of the most crucial steps in CoMFA and the results of CoMFA analyses depend on the alignment of molecules. Although the comparisons of different alignment techniques were reported [120, 121] 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. However, the alignment methods do not significantly affect the CoMSIA results. Therefore, only one alignment method was used in CoMSIA calculations.

4.1.4.1 The atom-based Root Means Square (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. The superimposition of both training and test sets based on RMS fit is shown in Figure 4.3 (a= 3'-processing and b = ST).

4.1.4.2 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. The superimposition of all compounds is illustrated in Figure 4.3 (c = 3'-processing and d = ST).

4.1.4.3 The rigid body field fit method.

In this method, the steric and electrostatic fields of the template were first calculated as reference fields, then the field fit procedure was used to change the orientation of all other molecule from their initial orientations in the first alignment such that their fields overlap the template fields as closely as possible, demonstrated in Figure 4.3 (e = 3'-processing and f = ST).

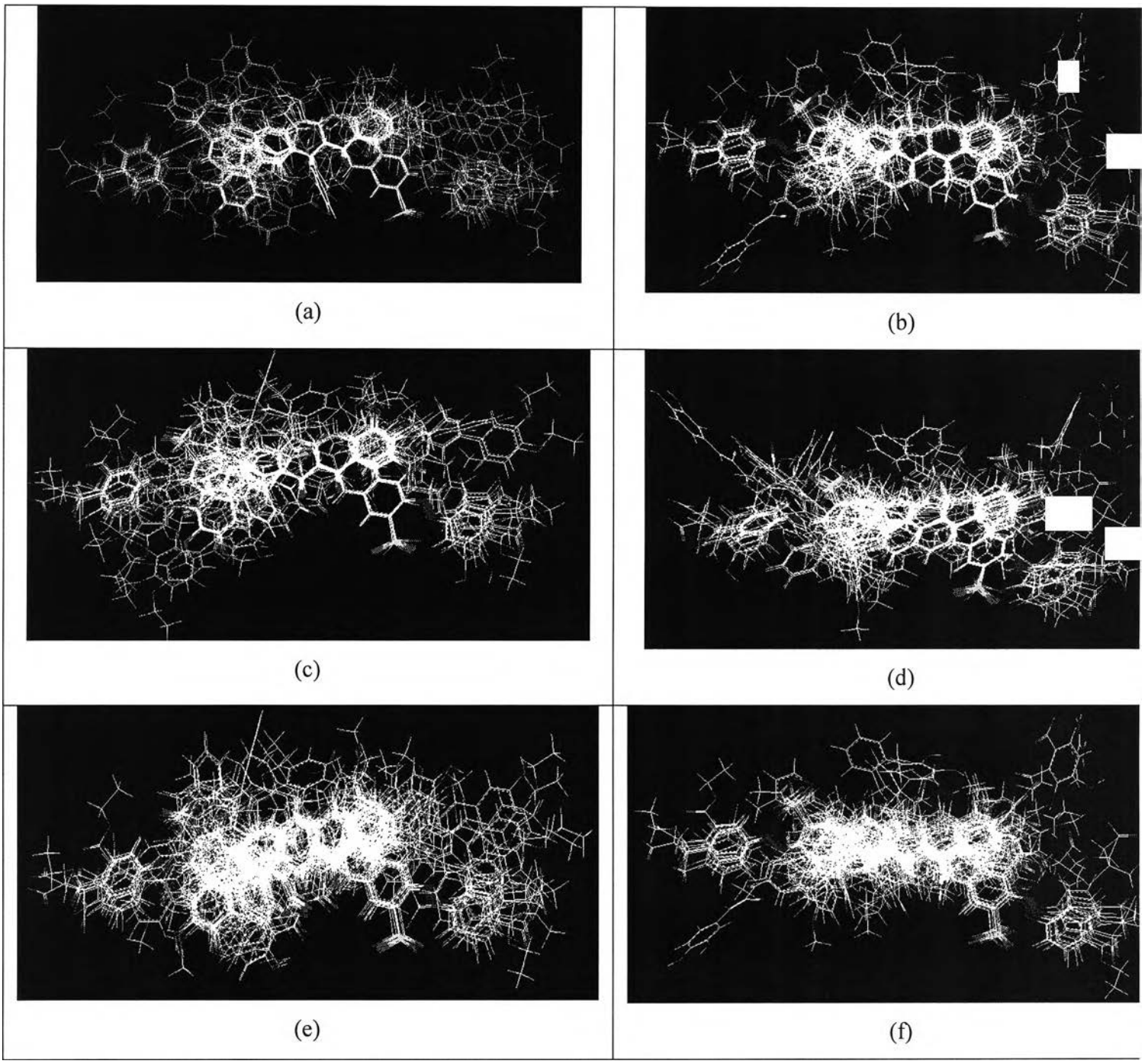


Figure 4.3 3D-view of aligned molecules (training and test sets) based on RMS fit (a = 3'-processing, b = ST), multi-fit (c = 3'-processing, d = ST), and field fit (e = 3'-processing, f = ST).

4.1.5 Field calculation for CoMFA

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 steric and electrostatic fields were calculated using a carbon sp3 probe atom with a charge of +1, and the energy cutoff was set to 30.0 kcal/mol. The partial atomic charges for each compound were assigned by the Gasteiger-Marsili method, implemented in the SYBYL 6.8. The analyses were performed with a scaling according to CoMFA standard deviations.

4.1.6 Field calculation for CoMSIA

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 was also used for CoMSIA field calculations. An sp3 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 (α=0.3) were used.

4.1.7 Region Focusing

Region focusing [122] is the application of weights to the lattice points in a CoMFA/CoMSIA region to enhance or attenuate the contribution of those points to subsequent analyses. It is usually applied in order to enhance the predictability of a 3D-QSAR study. The "StDev × Coefficient" values were used as weights, and different weighting factors were applied of which 0.5 was found to be most appropriate.

4.1.8 Derivation of models

After all CoMFA and CoMSIA fields were calculated, partial least square (PLS) analyses [123] were carried out to derive the models. Cross-validations [124, 125] in PLS were preformed by leave-one-out (LOO) procedure. The overall quality of the analyses was expressed by the cross-validated r^2 (r^2_{cv}) values. The optimum number of components (ONC), which yields the highest r^2_{cv} and the lowest standard error of prediction (SEP), were 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 CoMFA and CoMSIA results were interpreted graphically by field contribution maps using the field type "StDev ×Coeff".

4.2 Results and Discussion

4.2.1 3'-Processing activity

4.2.1.1 CoMFA Statistical Parameters

The statistical values of CoMFA results are summarized in Table 4.1. The CoMFA model using the atom-based RMS fit method has r^2_{cv} = 0.678 with 6 components, r^2 = 0.957 and r^2_{pred} = 0.719. The steric and electrostatic contributions are 70.5% and 29.5%, respectively. The flexible multi-fit alignment yields r^2_{cv} = 0.647 with 6 components, r^2 = 0.939 and r^2_{pred} = 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^2_{cv} of 0.546 using 6 components, r^2 = 0.953, and r^2_{pred} = 0.495, respectively. From this alignment, there are 70.6 % and 29.4 % for steric and electrostatic contributions, respectively. The steric and electrostatic contributions of all three models are almost similar (~70:30), indicating more important role of steric fields in ligand-receptor interactions. To improve the r^2_{cv}, 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^2_{cv} of this particular CoMFA model was increased from 0.678 to 0.698, whereas r^2 was 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^2_{cv} of all four models, the region focusing of the atom-based rms fit alignment with the highest r^2_{cv} was selected for generating CoMFA contour plots. The observed and predicted activities of the both training and test sets generated by a region focusing of atom-based RMS fit (Table 4.2) showed good consistency. The satisfactory quality of the best CoMFA model is represented in Figure 4.4, which shows a scatter plot of experimental pIC50 values versus the predicted pIC50 for the training set and the test set.

Table 4.1 Summary of CoMFA results of 3'-processing activity.

alignment	r^2_{cv}	ONC[d]	SEP[e]	r^2	SEE[f]	F value	r^2_{pred} [g]	Contributions	
								S [h]	E [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*]	**0.698**	**6**	**0.345**	**0.947**	**0.145**	**161.01**	**0.704**	**0.608**	**0.392**

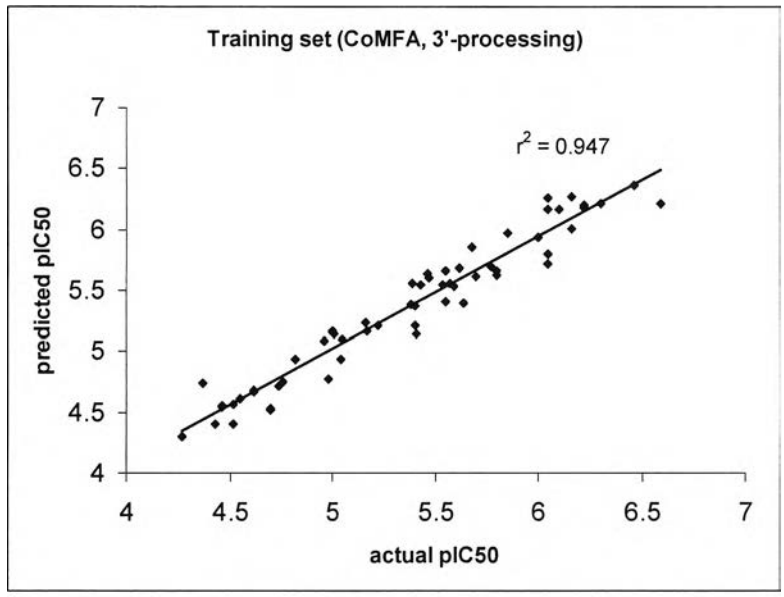
[a]Alignment by RMS fit; [b]Alignment by multi fit; [c]Alignment by field fit; [d]Optimum number of components; [e]Standard error of predictions; [f]Standard error of estimates; [g]Predictive r^2; [h] Steric; [i]Electrostatic; [*]Region focusing.

Table 4.2 Actual and calculated activities for 3'-processing mechanism of compounds used in training and test sets, obtained from CoMFA and CoMSIA.

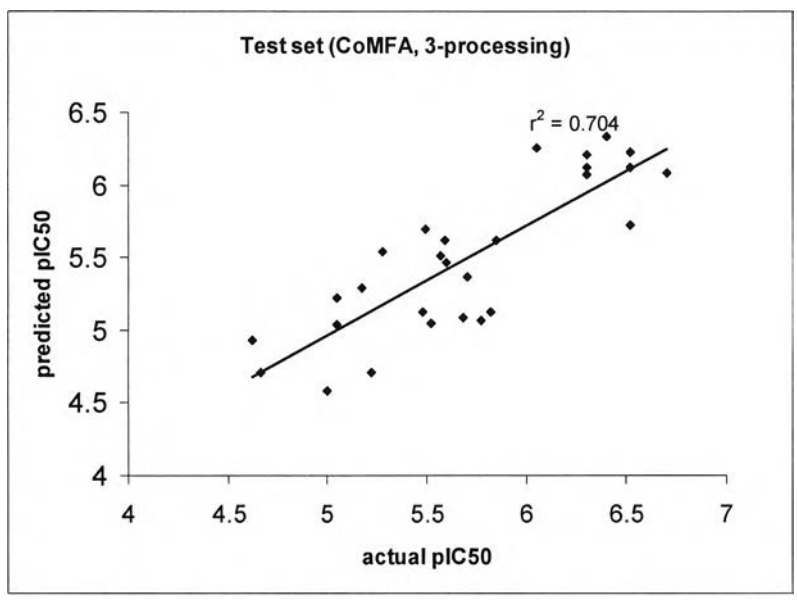
No.	3'-processing activity		
	Actual pIC_{50}	Predicted pIC_{50}	
		CoMFA	CoMSIA
1	4.76	4.75	4.74
2	5.16	5.24	5.15
3	5.16	5.23	5.12
4	5.00	5.16	5.01
5	4.82	4.94	4.97
6	4.70	4.53	4.71
7	4.46	4.54	4.68
8	4.37	4.74	4.70
9	4.27	4.30	4.76
10	4.62	4.68	4.92
11	5.41	5.14	4.99
12	5.59	5.53	5.28
13	5.54	5.55	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

No.	3'-processing activity		
	Actual pIC_{50}	Predicted pIC_{50}	
		CoMFA	CoMSIA
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
61	4.46	4.55	4.36
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

No.	3'-processing activity		
	Actual pIC_{50}	Predicted pIC_{50}	
		CoMFA	CoMSIA
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



(a)



(b)

Figure 4.4 Plot of observed and calculated activities against 3'-processing mechanism of (a) training set and (b) test set using CoMFA.

4.2.1.2 CoMFA Contour

CoMFA steric and electrostatic contours are shown in Figure 4.5 and 4.6, respectively. In order to investigate 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) [39]. 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 (Figure 4.5). 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**. 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 [126]. Three yellow contours are located around the tetrazole ring of 5CITEP suggesting that small bulky groups are required to increase 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 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 [127]. 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.

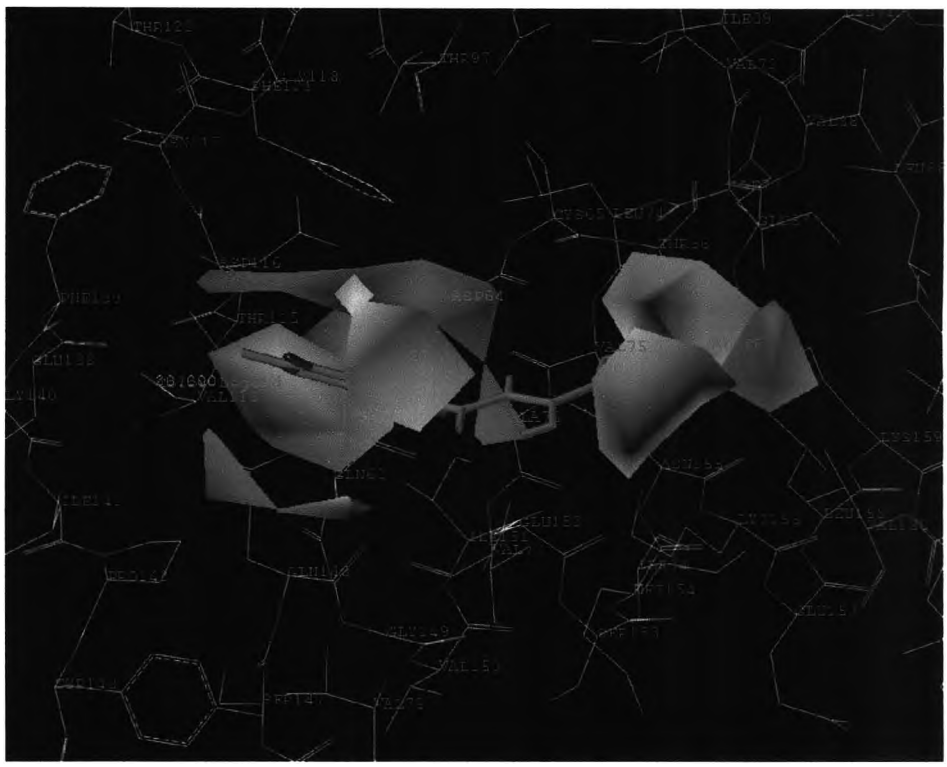


Figure 4.5 CoMFA StDev × coeff steric contour map for 3'-processing activity.

The CoMFA electrostatic contour plot is displayed in Figure 4.6. 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 proximity 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 NO_2 and the C_6H_5COO substituents of compound **9** and compound **12**, respectively, are located in this blue region. The NO_2 substituent is a strong electron withdrawing group, whereas C_6H_5COO 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 group carbon of 5CITEP, indicating a preference for negatively charged substituents in these areas. As expected, red contours are close to the Mg^{2+}, which is located between Asp64 and Asp116 of HIV-1 IN. The aromatic moiety of IN inhibitors such as chicoric acids, has been proposed to interact with this divalent cation via a cation-π type interaction [110, 115], or by a charge-charge interaction between metal ion and partial or ionic charges of inhibitors [115]. Therefore, substituents with high electron density will strongly interact with Mg^{2+} leading to an enhanced biological activity of HIV-1 IN inhibitors. Additionally, the presence of negative charge favored red contours surrounding the tetrazole ring indicates that electron rich groups may increase binding affinity. These contours were observed near residues Asn155, Lys156 and Lys159 which favor a negative charge.

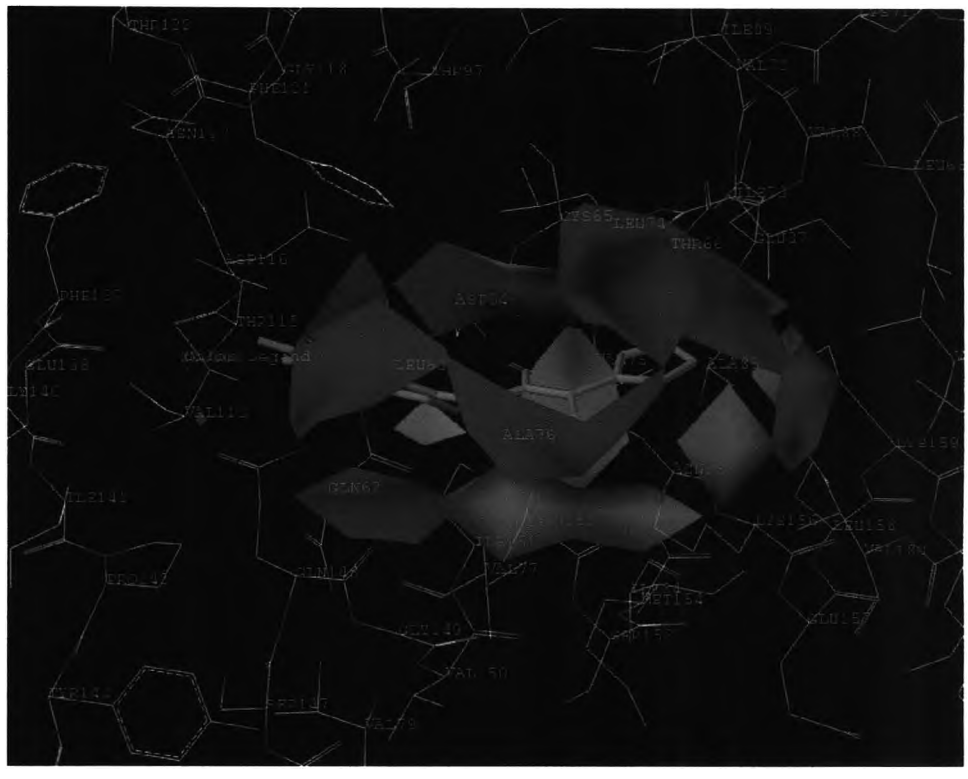


Figure 4.6. CoMFA StDev × coeff electrostatic contour map for 3'-processing activity.

4.2.1.3 CoMSIA Statistical Parameter

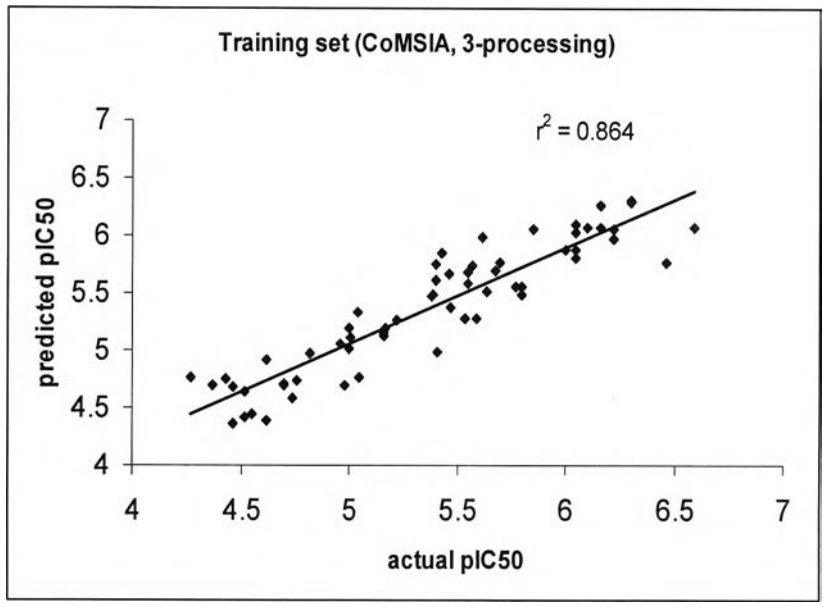
A total of 16 CoMSIA models were generated using either single or combined fields. The alignment giving the highest r^2_{cv} in CoMFA, atom-based RMS fit method, was used. The statistical parameters are summarized in Table 4.3. Among the different field combinations, the CoMSIA model number 11 with three fields, steric, hydrogen bond donor and hydrogen bond acceptor, gave the highest statistical values, i.e. r^2_{cv} = 0.693, ONC =5 , r^2 = 0.872 and r^2_{pred} = 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 (r^2_{cv}) and external predictions (r^2_{pred}) (except the CoMSIA model number 2 which was derived from only the electrostatic field). The region

focusing was applied to CoMSIA model number 11. After region focusing (model number 12), the r^2_{cv} of this specific model was increased from 0.693 to 0.724 (ONC = 5). The r^2 = 0.864 and r^2_{pred} = 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 [128], 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 4.2 and Figure 4.7. These demonstrate a good performance of CoMSIA model.

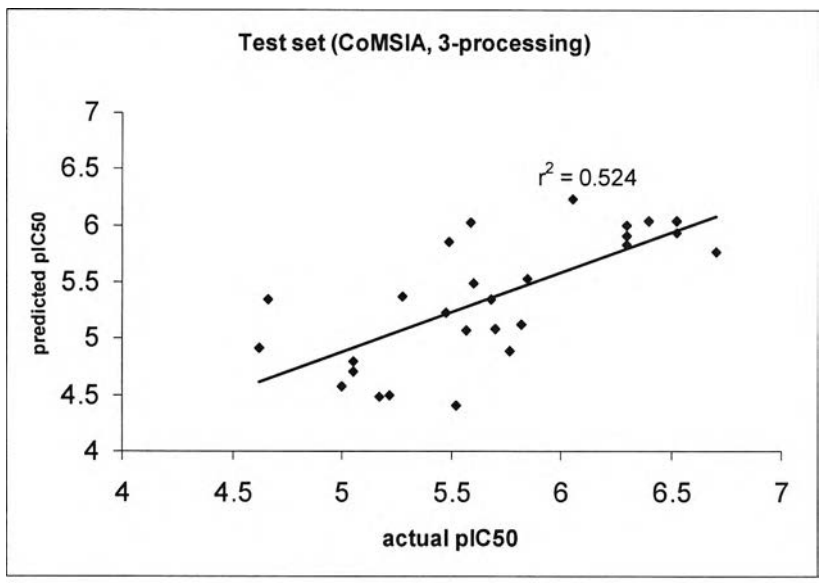
Table 4.3 Summary of CoMSIA results of 3'-processing activity.

model	fields	r^2_{cv}	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	nd	1.000	-	-	-	-
2	E	0.004	1	0.600	0.207	0.535	15.37	nd	-	1.000	-	-	-
3	H	0.426	5	0.392	0.846	0.244	60.52	nd	-	-	1.000	-	-
4	D	0.611	6	0.501	0.794	0.285	34.69	nd	-	-	-	1.000	-
5	A	0.330	3	0.471	0.601	0.386	28.59	nd	-	-	-	-	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***	**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; *Region focusing; nd = not determine.



(a)



(b)

Figure 4.7 Plot of observed and calculated activities against 3'-processing mechanism of (a) training and (b) test sets using CoMSIA.

4.2.1.4 CoMSIA Contour

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 (Figure 4.8), 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) [39].

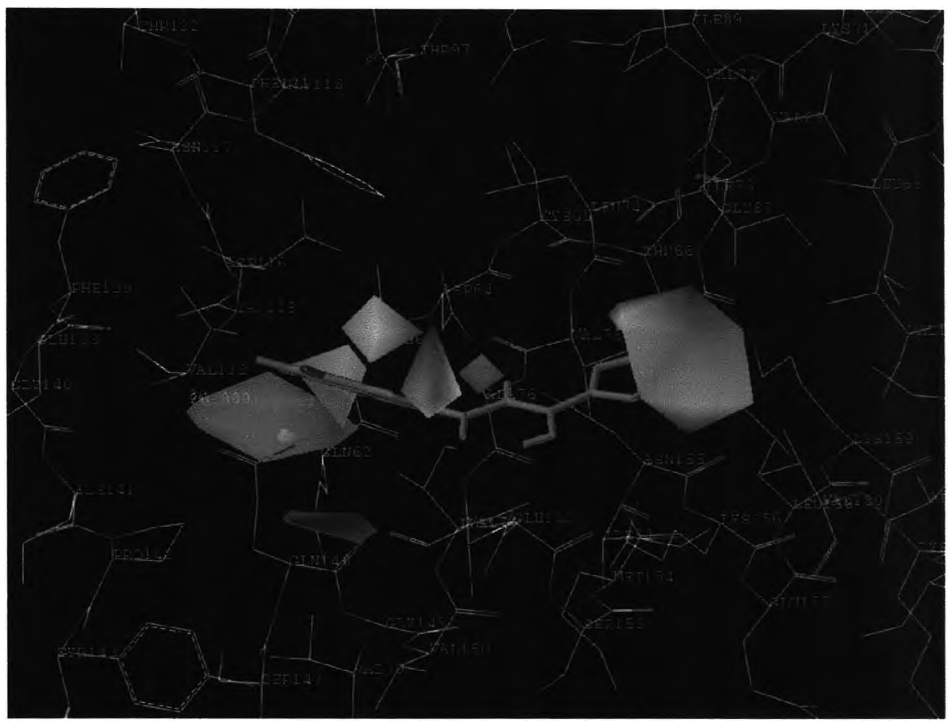


Figure 4.8 CoMSIA StDev × coeff steric contour map for 3'-processing activity

Figure 4.9 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 a NH moiety located near this cyan contour. This cyan contour correponds 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 [129]. 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 [101].

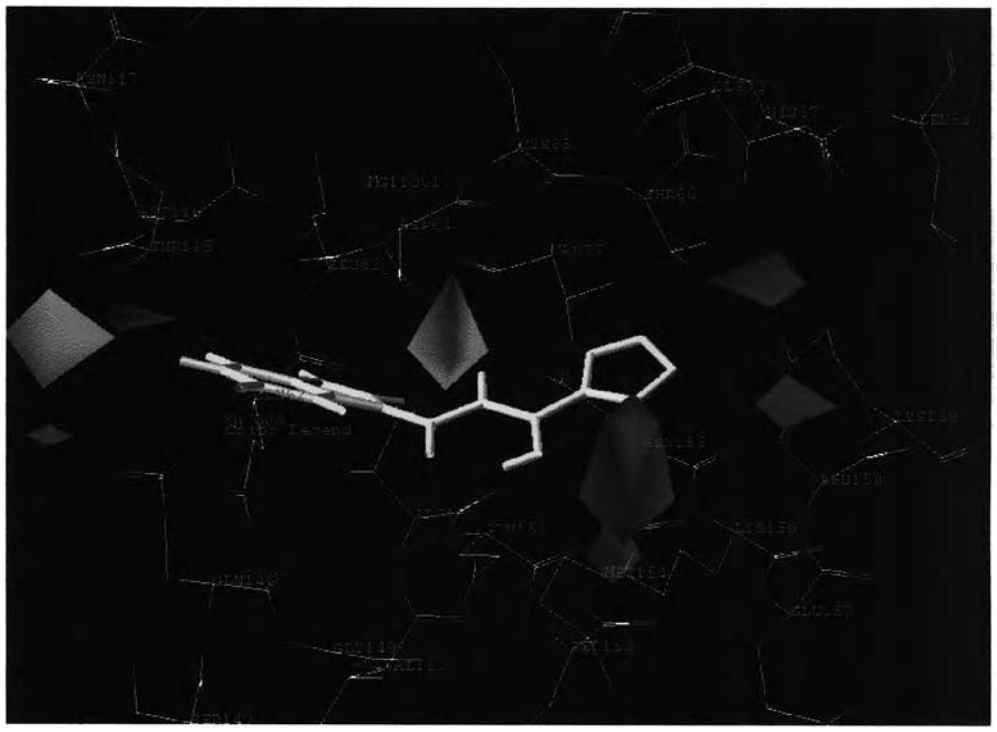


Figure 4.9 Mapping of CoMSIA StDev × coeff hydrogen bond donor contour plots within the active site of the complex structure of HIV-1 IN/5CITEP for 3'-processing activity.

Figure 4.10 shows the CoMSIA hydrogen bond acceptor field, denoted by magenta and red contours. Magenta contours 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 **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 [130]. 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 C=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.

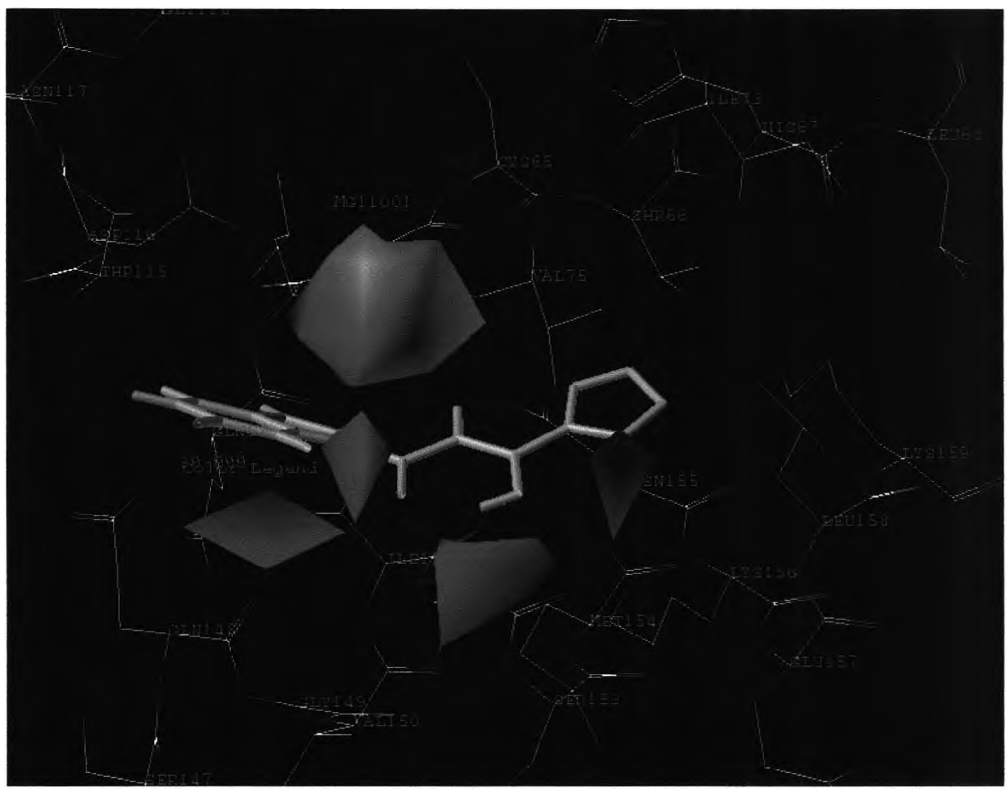


Figure 4.10 Mapping of CoMSIA StDev × coeff hydrogen bond acceptor contour plots within the active site of the complex structure of HIV-1 IN/5CITEP for 3'-processing activity.

4.2.2 ST activity

4.2.2.1 CoMFA Statistical Parameter

CoMFA statistical analyses are summarized in Table 4.4. The CoMFA model with atom-based RMS fit alignment gave r^2_{cv} = 0.625 with 5 components, r^2 = 0.920 and

r^2_{pred} = 0.650. The steric and electrostatic contributions are 63.6% and 36.4%, respectively. Using the multi-fit alignment method, CoMFA yielded r^2_{cv} = 0.672, 5 components, r^2 = 0.915 and r^2_{pred} 0.552. Steric and electrostatic contributions for this model were 62.7% and 37.3 %, respectively. CoMFA model using rigid body field fit showed poor r^2_{cv} = 0.524 with 4 components, r^2 0.849, and r^2_{pred} =0.283. From this alignment procedure, steric and electrostatic contributions were 65.6 % and 34.4 %, respectively. Likewise 3'-processing, all of the three CoMFA models show a similar high contribution of steric (~ 60%) than that of electrostatic field (~ 40%). This means that the predominant of steric interactions is required to stabilize the protein-ligand complex. The multi-fit alignment method which shows the highest r^2_{cv} was further used to perform region focusing. After region focusing, the r^2_{cv} significantly increased from 0.672 to 0.720. The r^2 and r^2_{pred} were 0.926 and 0.570 respectively. The region focusing of the multi-fit alignment was selected to generate the contour plots. The values of the observed and calculated activities of training and test sets are shown in Table 4.5 and the graph between the actual and predicted pIC50 values is shown in Figure 4.11, indicating a good statistical correlation between estimated and actual biological activities.

Table 4.4 CoMFA statistical parameters for ST activity.

alignment	r^2_{cv}	ONC[d]	SEP[e]	r^2	SEE[f]	F value	r^2_{pred} [g]	Contributions	
								S [h]	E [i]
1[a]	0.625	5	0.492	0.920	0.228	132.81	0.650	0.636	0.364
2[b]	0.672	5	0.460	0.915	0.234	124.51	0.552	0.627	0.373
3[c]	0.524	4	0.549	0.849	0.309	83.20	0.283	0.656	0.344
2[a*]	**0.720**	**6**	**0.428**	**0.926**	**0.221**	**118.131**	**0.570**	**0.560**	**0.440**

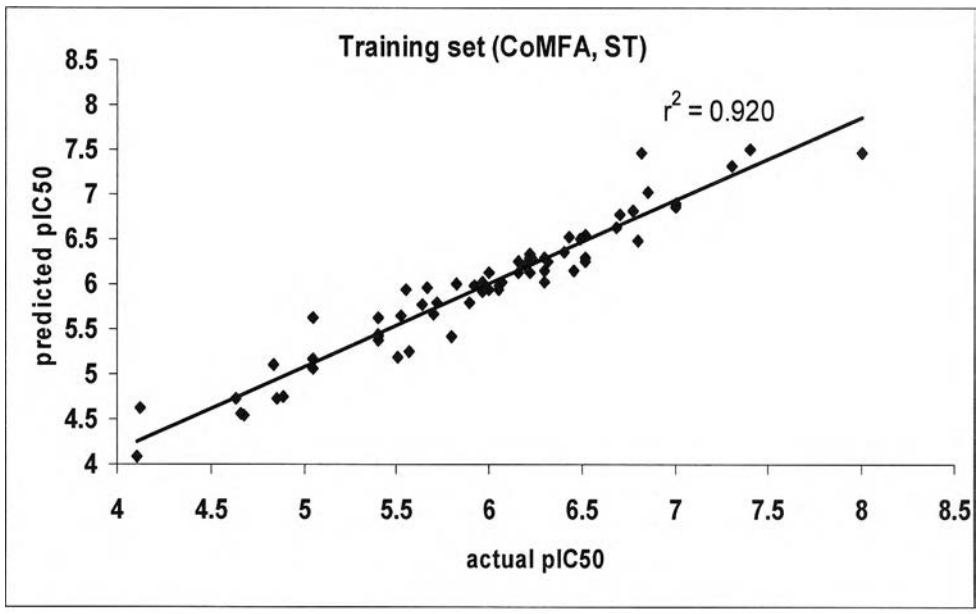
[a]Alignment by RMS fit; [b]Alignment by multi fit; [c]Alignment by field fit; [d]Optimum number of components; [e]Standard error of predictions; [f]Standard error of estimates; [g]Predictive r^2; [h] Steric; [i]Electrostatic; [*]Region focusing

Table 4.5 Actual and calculated activities for ST mechanism of compounds used in training and test sets, obtained from CoMFA and CoMSIA.

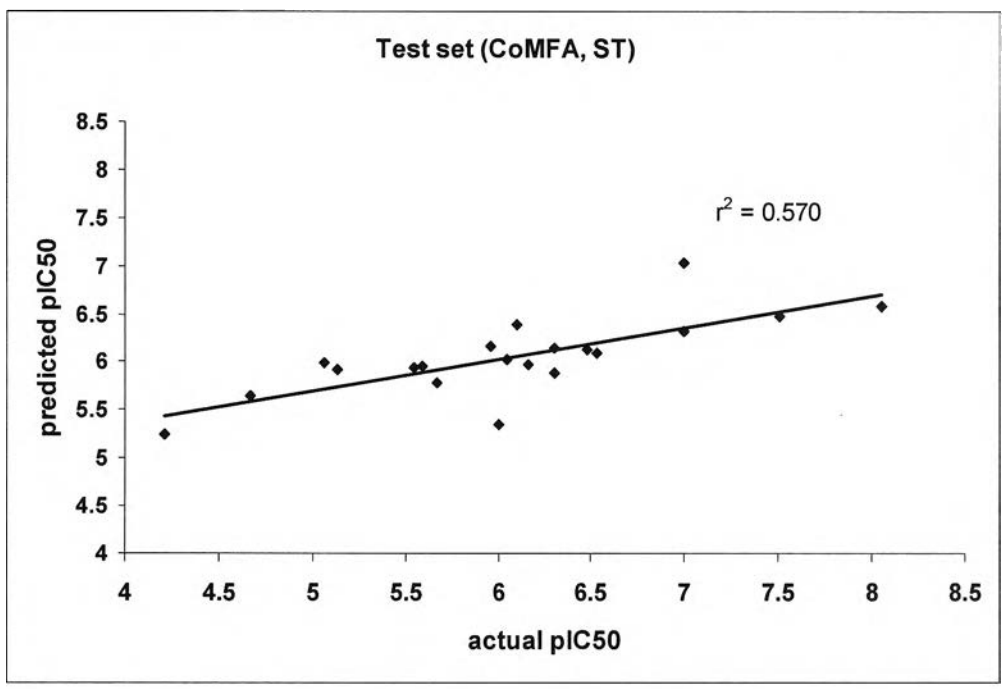
No.	ST activity		
	Actual pIC_{50}	Predicted pIC_{50}	
		CoMFA	CoMSIA
1	4.89	4.75	4.62
2	4.12	4.62	4.55
3	4.66	4.56	4.53
4	4.68	4.54	4.52
5	6.32	6.26	6.50
6	6.69	6.63	6.24
7	6.49	6.50	6.28
8	5.82	6.01	6.06
9	6.46	6.14	6.40
10	7.00	6.86	7.01
11	6.82	7.45	7.28
12	6.85	7.02	7.15
13	7.00	6.89	7.13
14	8.00	7.45	7.18
15	7.40	7.50	7.35
16	6.43	6.53	6.69

No.	ST activity		
	Actual pIC_{50}	Predicted pIC_{50}	
		CoMFA	CoMSIA
17	7.30	7.32	7.28
18	4.84	5.10	5.06
19	5.57	5.24	5.44
20	4.11	4.08	4.48
21	4.63	4.72	4.53
22	5.96	6.02	5.69
23	6.24	6.28	6.46
24	6.40	6.35	6.38
25	5.52	5.65	5.38
26	6.07	6.02	6.00
27	5.72	5.80	5.65
28	6.80	6.48	6.67
29	6.77	6.82	6.54
30	5.92	5.97	5.62
31	5.51	5.18	5.32
32	5.05	5.16	5.26
33	5.40	5.37	5.66
34	5.40	5.44	5.62
35	6.30	6.29	6.05
36	6.16	6.26	6.24
37	6.16	6.12	6.21
38	6.30	6.15	6.00
39	6.00	5.93	6.04
40	6.16	6.12	6.05
41	5.40	5.62	5.83
42	5.89	5.79	6.04
43	6.22	6.33	5.99
44	5.05	5.07	5.11
45	6.05	5.94	6.05
46	5.96	5.92	6.07
47	5.66	5.95	5.90
48	5.64	5.78	6.01
49	5.05	5.63	4.98
50	6.30	6.02	5.83
51	6.05	5.98	6.21
52	5.40	5.41	5.69
53	5.80	5.42	5.52
54	6.52	6.24	6.28
55	6.52	6.55	6.05
56	6.70	6.78	6.67
57	6.22	6.30	5.91
58	6.22	6.13	6.25

No.	ST activity		
	Actual pIC_{50}	Predicted pIC_{50}	
		CoMFA	CoMSIA
59	5.70	5.67	5.97
60	4.85	4.73	4.79
61	6.00	6.13	6.29
62	5.55	5.93	5.83
63	6.52	6.29	6.70
64	6.19	6.20	6.24
65	4.67	5.65	4.76
66	5.67	5.78	5.03
67	7.00	6.32	6.48
68	7.00	7.03	7.08
69	8.05	6.58	7.79
70	7.51	6.47	6.11
71	5.06	5.99	5.17
72	6.53	6.09	5.58
73	4.21	5.25	5.46
74	6.1	6.39	6.45
75	6.48	6.13	6.20
76	6.16	5.98	5.98
77	6.30	5.88	5.35
78	5.55	5.94	6.08
79	5.59	5.96	5.84
80	6.05	6.03	5.83
81	5.13	5.92	5.90
82	6.30	6.15	6.12
83	6.00	5.35	4.89
84	5.96	6.17	6.82



(a)



(b)

Figure 4.11 Plot of observed and calculated activities against ST mechanism of (a) training and (b) test sets using CoMFA.

4.2.2.2 CoMFA contours

The CoMFA contours plots of steric and electrostatic interactions mapped to the X-ray co-crystal structure of HIV-1 IN-5CITEP are displayed in Figure 4.12 and 4.13, respectively.

It is apparent that the pattern of steric field of ST process is somewhat similar to that of 3'-processsing. Sterically favored green area is noticed around chloroindole ring while sterically disfavored yellow region is observed near to tetrazole ring of 5CITEP. In Figure 4.12, the big green contour beyond the chloroindole ring of 5CITEP suggests that more bulkier group in this area is expected to enhance the activity. This conclusion is consistent well with the experiment data that the DKA derivatives (compounds **5-17 and 67-70**) are generally more active than 5CITEP (except compound **8**). Considering DKA analogue (compounds **5-17** and **67-70**), compound **8** has less activity than the others as it does not contain any substitution pointing toward this green area. In addition, the low inhibitory potencies of catechol (compounds **1-4** and **65-66**) are probably because this class of compounds has no functional group extended to this area. This steric favorable region could be related to the hydrophobic or π-π stacking interaction between aliphatic or aromatic parts of ligands and hydrophobic amino acid residues Ile141, Pro142 and Tyr143 as aforementioned in the case of 3'-processing. The other two green contours were noticed at the Cl and close to the N of chloroindole ring of 5CITEP. The latter contour can be used to explain the activities of cyclohexanone analogues (compounds **44-56** and **76-77**). Compounds **54, 55,** and **56** oriented the $COOCH_2CH_3$, COO^- and CH_2COO^-, respectively, attached at their central aliphatic ring toward to this green zone, thus, these 3 compounds exhibit higher potency than other compounds (compounds **44-53** and **76-77**) in the same group. On the other hand, compound **44** has no functional group pointing to this position, it therefore shows poor activity. Moreover, although compound **47** has methylphenyl substitution at the central aliphatic ring, this particularly part of ligand is quite far from this green area. Thus, it shows the lower activity in comparison with compound **56**.

Two yellow contours were observed near the chloroindole ring of 5CITEP implying that bulky substituents there may decrease the inhibitor activity of ligands. However, these sterically disfavored regions were positioned close to steric favored green

area implying that the size of the substituent should not be too large and could have an appropriate size extending to this position. Another large yellow contour was noticed near the tetrazole ring of 5CITEP. The aromatic sulfonamides containing compounds (compounds 18-21) generally exhibit activities lower than DKA because their phenylsulfone substituents embedded this yellow region. In the mean time, this observation explained why the activities of carbonyl J derivatives (compounds 34-43) whose naphthalene solfonic acid substituent overlapped this sterically disfavorable area are slightly lower than DKA analogues. Likewise 3'-processing, amino acid residues Asn155, Lys156 and Lys159 were located close to this sterically disfavored contour.

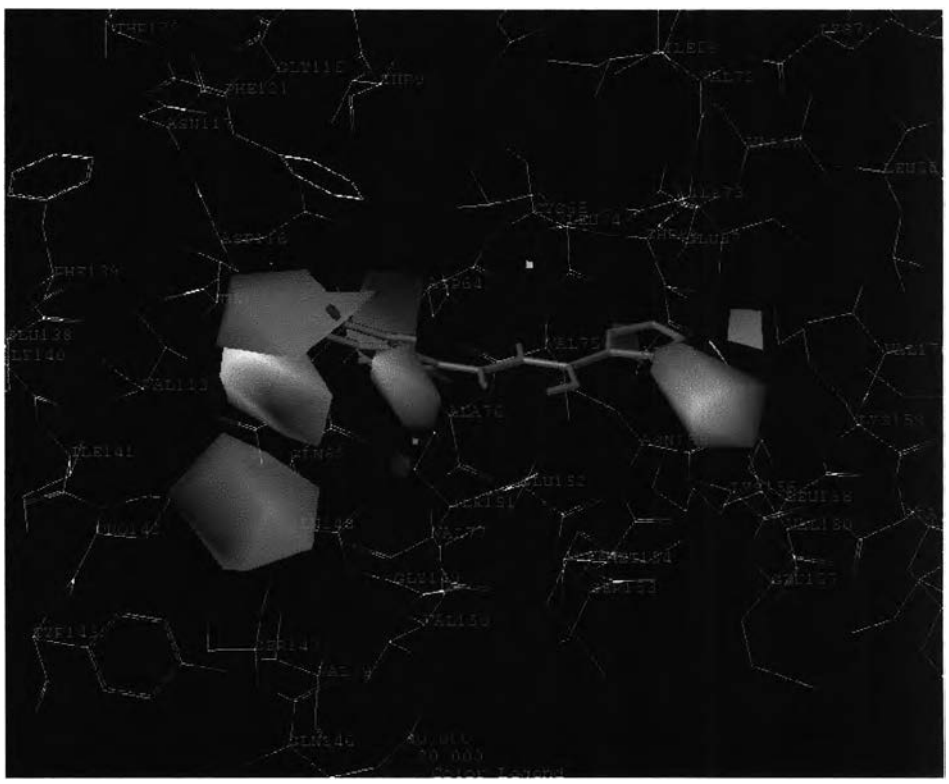


Figure 4.12 CoMFA StDev × coeff steric contour map for ST activity.

Electrostatic contours merged to the HIV-1 IN-5CITEP complex are shown in Figure 4.13. Large blue areas are spread above the plane of 5CITEP. This requirement of electropositive charge group explained why many of DKA analogues generally exhibit higher potency than salicylhydrazine (compounds **57-59**). The central pyrozolone ring, an electron rich moiety, of salicylhydrazine is located in this blue area while no high electron density functions of DKA was found in this blue zone. Moreover, the NO_2, OH, and OCH_3 substituents in the benzene ring of compounds **57** and **58**, are in the vicinity of the small blue contours in the left hand side so they exhibit low inhibitory potency. Additional large blue areas were found near the keto-enol moiety of 5CITEP and suggested that low electron density in this area is expected to increase the activity. Glu152, a negatively charged amino acid, is located close to this contour. Thus, substituents with low electron density in this region are required as they may stronger interact with the carboxylate of Glu152. The other blue zones were noticed beyond the terazole ring of 5CITEP. This profile explains the poor activity of compounds **18-21** in comparison with DKA as they orient the phenylsulfone, the high electron density moiety, toward this location.

The presence of several red contours around tetrazole ring of 5CITEP indicates that the electron rich groups would be beneficial to the binding affinity. Compounds with higher HIV-1 IN inhibitory activities, for instance, DKA analogues, generally bear a high electron density group, i.e. carboxylate in this position. Similar to the case of 3'-processing, positive charge residues Lys156 and Lys159 locate close to this red region. Therefore, an increase in negative charge of ligands would increase activity and supports the better interaction of inhibitors with electron deficient amino acids of HIV-1 IN.

Figure 4.13 CoMFA StDev × coeff electrostatic contour map for ST activity.

4.2.2.3 CoMSIA Statistical Parameter

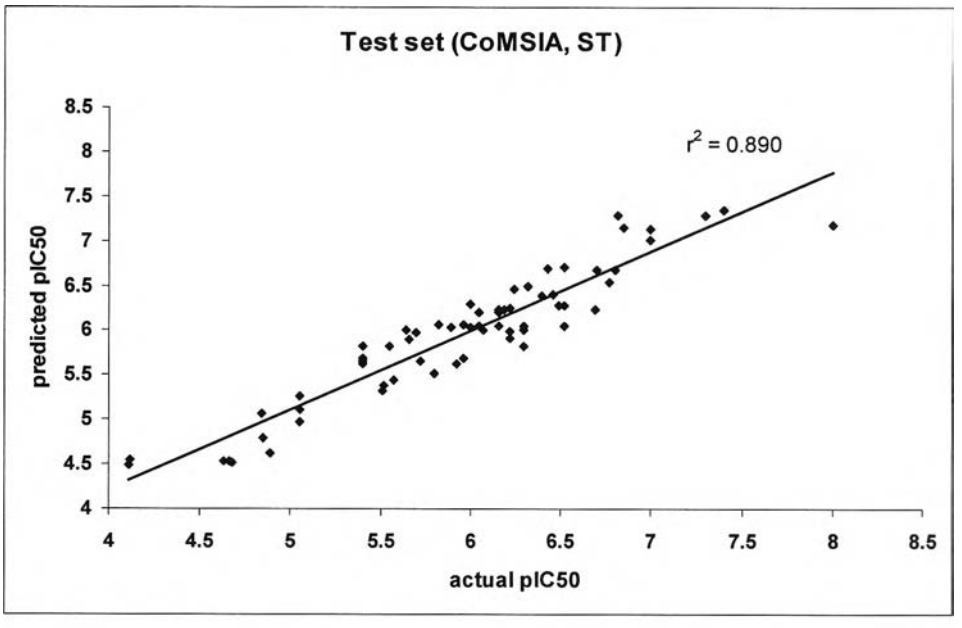
Table 4.6 lists the statistical parameters of CoMSIA models using multi-fit alignment method, the highest r^2_{cv} in CoMFA. As in the 3'-processing, the CoMSIA model number 11 with steric, hydrogen bond donor and hydrogen acceptor fields showed the highest r^2_{cv} = 0.591 with 3 ONC, r^2 = 0.829 and r^2_{pred} = 0.445. The steric, hydrogen bond donor and hydrogen bond acceptor contributions were 24.9%, 37.1% and 37.9%, respectively, indicating that steric field together with the hydrogen bond donor and the hydrogen bond acceptor interactions in protein-ligand complex. The other CoMSIA models also show acceptable internal and external predictive ability (except CoMSIA model number 2 which was generated by electrostatic field only). The CoMSIA model number 11 was further performed region focusing. After region focusing (model 12), the r^2_{cv}, ONC, r^2, and r^2_{pred} was 0.656, 5, 0.890, and 0.461, respectively. Steric, hydrogen bond donor and hydrogen bond acceptor contributions are 34.2%, 31.5% and 34.1%,

respectively, indicating that steric fields together with the hydrogen bond donor and hydrogen bond acceptor fields interaction equally dominate the binding between HIV-1 IN and its inhibitors. The experimental versus predicted biological activity are given in Table 4.5. The plot of prediction of training and test set as depicted in Figure 4.14 indicates an acceptable CoMSIA model.

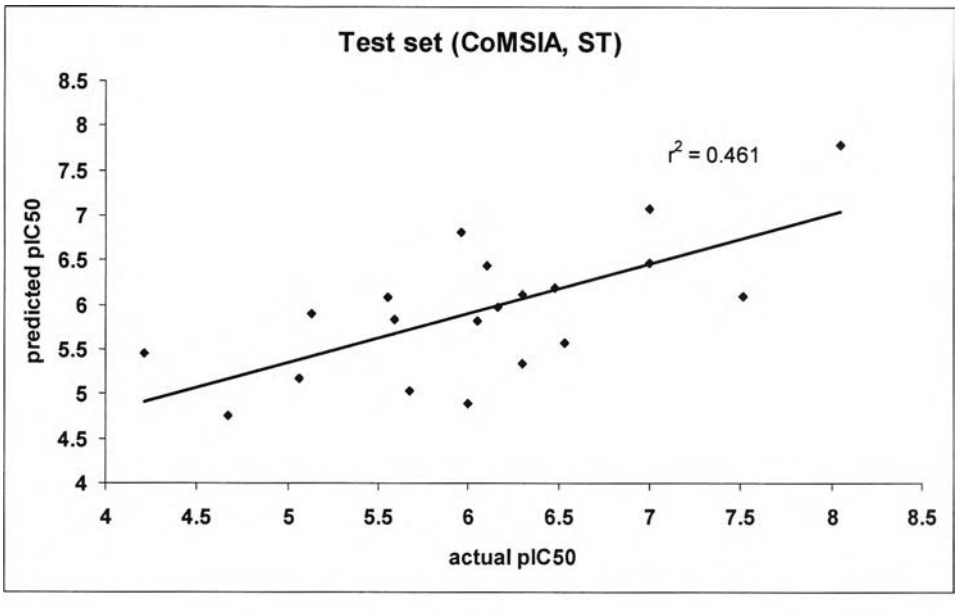
Table 4.6 CoMSIA statistical parameters for ST mechanism.

model	fields	r^2_{cv}	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	0.555	5	0.536	0.842	0.320	61.70	nd	1.000	-	-	-	-
2	E	-0.001	2	0.784	0.368	0.622	17.79	nd	-	1.000	-	-	-
3	H	0.448	6	0.602	0.875	0.286	66.70	nd	-	-	1.000	-	-
4	D	0.489	6	0.579	0.783	0.378	34.21	nd	-	-	-	1.000	-
5	A	0.364	2	0.625	0.581	0.507	42.30	nd	-	-	-	-	1.000
6	S+E	0.578	6	0.526	0.874	0.288	65.62	0.376	0.484	0.516			
7	S+E+H	0.488	6	0.580	0.908	0.246	93.69	0.395	0.273	0.351	0.375	-	-
8	S+E+D	0.457	6	0.597	0.920	0.229	109.63	0.412	0.269	0.328	-	0.403	-
9	S+E+A	0.490	5	0.490	0.889	0.267	93.17	0.302	0.255	0.359	-	-	0.386
10	D+A	0.492	3	0.563	0.752	0.394	60.53	0.380	-	-	-	0.499	0.501
11	D+A+S	0.591	3	0.505	0.829	0.326	97.29	0.445	0.249	-	-	0.371	0.379
12	**D+A+S***	**0.656**	**5**	**0.471**	**0.890**	**0.266**	**93.79**	**0.461**	**0.342**	**-**	**-**	**0.315**	**0.343**
13	D+A+E	0.381	5	0.632	0.861	0.299	72.14	0.392	-	0.355	-	0.327	0.318
14	D+A+H	0.577	5	0.522	0.897	0.258	101.16	0.412	-	-	0.373	0.318	0.308
15	S+E+D+A	0.486	5	0.576	0.902	0.252	106.35	0.388	0.183	0.281	-	0.272	0.264
16	All	0.539	6	0.550	0.934	0.208	134.86	0.400	0.125	0.217	0.230	0.219	0.209

[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; *Region focusing. nd = not determine.



(a)



(b)

Figure 4.14 Plot of observed and calculated activities against ST mechanism of (a) training and (b) test sets using CoMSIA.

4.2.2.4 CoMSIA Contour

The CoMSIA model number 12 was used to produce the contour maps. As CoMSIA steric contour plots share similar pattern to that of CoMFA, only hydrogen bond donor and acceptor fields are shown. To gain better understanding of the interaction between enzyme and inhibitor, amino acids surrounding 5CITEP in the binding pocket were merged into the CoMSIA contours (Figure 4.15).

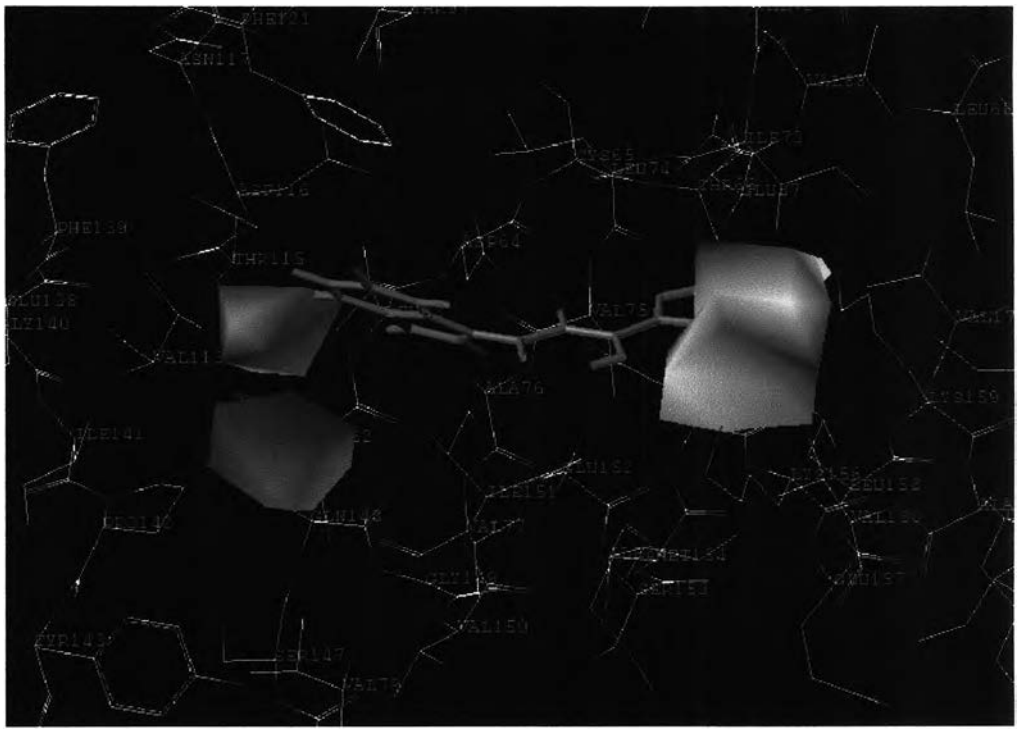


Figure 4.15 CoMSIA StDev × coeff steric contour map for ST activity.

CoMSIA hydrogen bond donor plots are depicted in Figure 4.16. Two cyan contours were observed. Similar to 3'-processing results, one cyan maps is found near the CH adjoining to the carbonyl carbon and to the hydroxyl carbon of 5CITEP. This suggests that hydrogen bond donor property of ligand is required as it can form hydrogen bonding interaction with the carbonyl oxygen of Asp64 which is in the proximity of the cyan contour. Another cyan contour was found beyond the tetrazole ring of 5CITEP. It can be used to explain the trends of biological data of cyclohexanone derivatives. The

inhibitory potency of compound **56** is higher than that of compound **44,** because the hydroxyl substituents of the benzene ring of the former compound oriented toward this cyan area. On the contrary, compound **44** has no hydrogen bond donor group, hence its inhibitory activity is very poor. Two purple contours were noticed at the opposite site, i.e. the chloroindole and tetrazole ring, of 5CITEP. The low inhibitory potency of curcumin (compounds **31-33**) relative to DKA analogues is due to the existing of the OH of catechol moiety embedded in these areas. Moreover, the catechol group of compounds **1-4** oriented toward the purple region close to tetrazole ring leading to their low inhibitory activities.

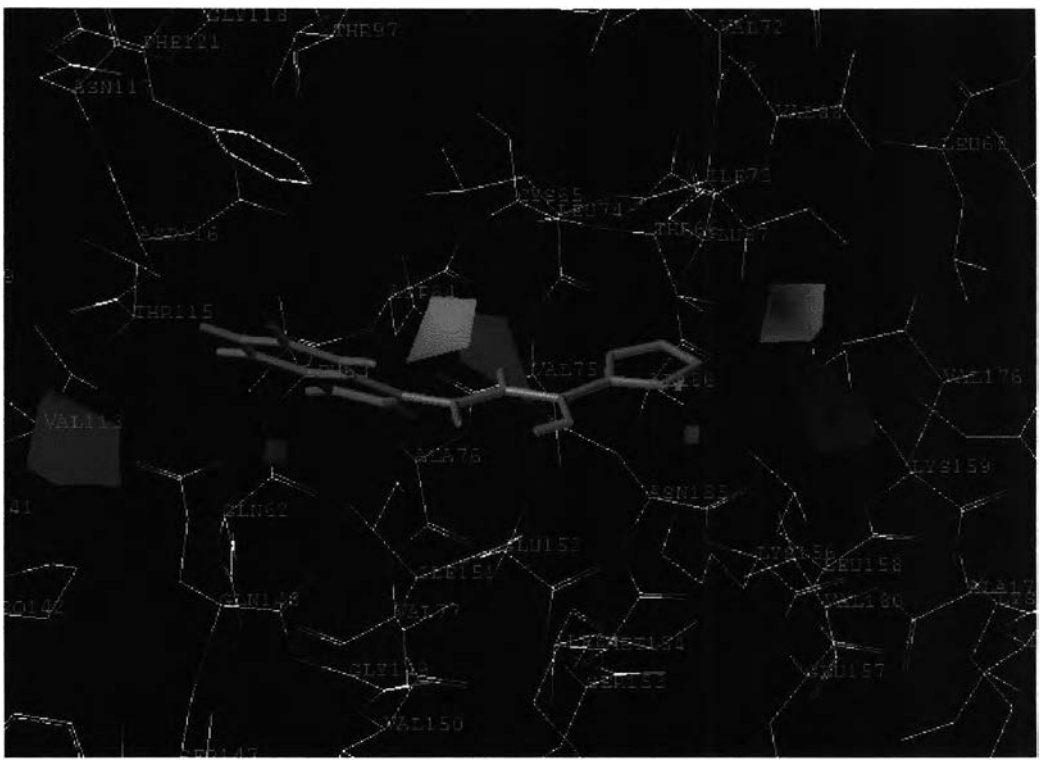


Figure 4.16 Mapping of CoMSIA StDev*coeff hydrogen bond donor contour plots within the active site of the complex structure of HIV-1 IN/5CITEP for ST activity.

CoMSIA hydrogen bond acceptor field illustrated in Figure 4.17 is quite similar to that of 3'-processing. Only one magenta contour is present in the vicinity of keto-enol group of 5CITEP. This indicates that the presence of hydrogen bond acceptor groups of

ligand is preferable in this position. The high inhibitory activity of DKA (compounds **5-14**) or naphthyridines derivatives (compounds **15-17**) is because the presence of their carbonyl oxygen near to this magenta map. Meanwhile, there is no hydrogen bond acceptor group of catechol (compounds **1-4**) or phenylsulfone containing compounds (compounds **18-21**) in this area, therefore, this class of inhibitors show lower biological activities in comparison with the DKA or DKA-like compounds. The Glu152 which locates near to the magenta contour may form hydrogen bond with the acceptor group of ligands, thereby, enhances the binding between them. The large red zone near the NH of indole ring of 5CITEP indicates that hydrogen bond acceptor groups on ligand should not be presented. This is well consistent with the biological data of tetracyclines (compounds **60-63**). Among this group of compounds, compound **60,** has the worse activity as its N=N, hydrogen acceptor functionality, is positioned in this red area. On the other hand, compound **63**, the most potent inhibitor of this class, has no hydrogen bond acceptor group in this location. Residue Asp64 found near to this contour can act as an acceptor to form a hydrogen bond with the donor group of ligands and this increases the binding affinity.

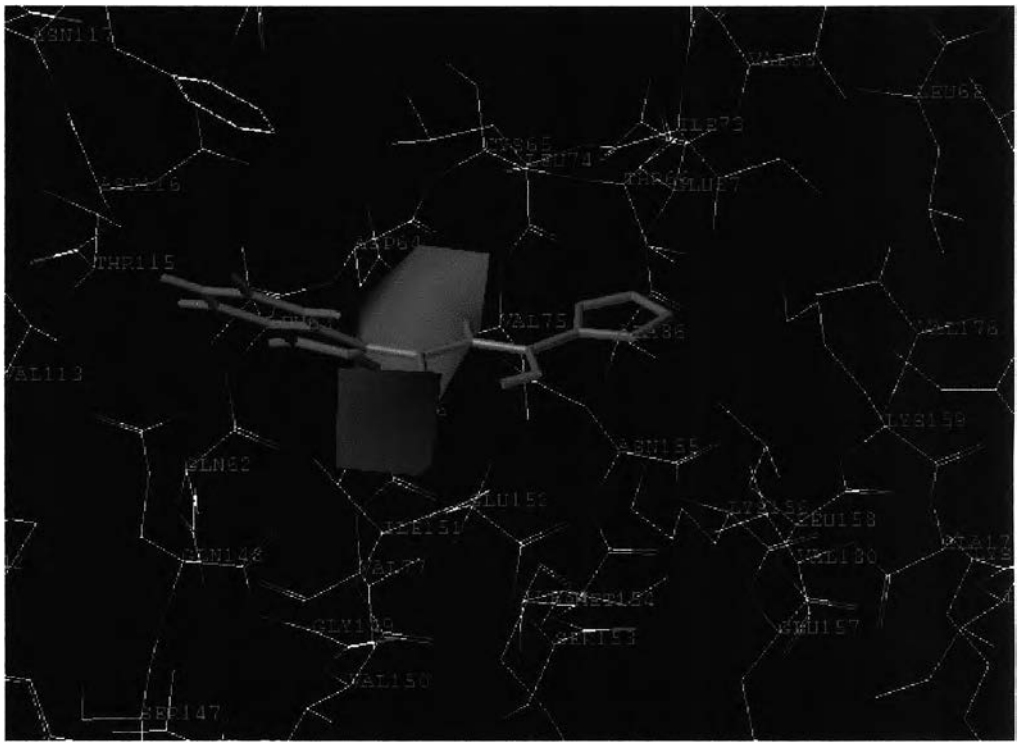


Figure 4.17 Mapping of CoMSIA StDev*coeff hydrogen bond acceptor contour plots within the active site of the complex structure of HIV-1 IN/5CITEP for ST activity.

4.3 Conclusion

In this study, a single CoMFA model and a single CoMSIA model that can well describe activities of diverse structural classes for each 3'-processing (89 compounds) and ST (84 compounds) activities were successfully derived. 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 generated from the field fit method gave a poor r^2_{cv} (3'-processing =0.546, ST =0.524). The alignment method showing the highest r^2_{cv} of each activity (RMS fit for 3'-processing and multi fit for ST activities) was further used to derive CoMSIA model. For compounds inhibiting 3'-processing reaction of HIV-1 IN, the best CoMFA and CoMSIA models show r^2_{cv} of 0.698 and of 0.724, respectively, which indicate a good correlation between predicted and observed inhibitory potencies.

For compounds ST activities, the highest r^2_{cv} of 0.720 and 0.656 were obtained from CoMFA and CoMSIA, respectively. The CoMFA results of both reactions suggest the more requirement of steric field (~ 60%) than electrostatic field (~ 40%) on ligand-receptor interaction.

The CoMFA contour maps of 3'-processing and ST inhibitory activities are somewhat similar to each other. From the contour analyses, 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, functional groups with high electron density in the ligand are necessary for a better interaction with the metal ion or the positive charge amino acids such as Lys156 and Lys159 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 were 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 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.