



## 3D-QSAR and molecular modeling of HIV-1 integrase inhibitors

Mahindra T. Makhija & Vithal M. Kulkarni\*

Pharmaceutical Division, Department of Chemical Technology, University of Mumbai, Matunga; Mumbai 400 019 India

Received 24 October 2001; accepted 1 May 2002

**Key words:** comparative molecular field analysis, comparative molecular similarity indices analysis, docking, HIV-1 integrase, partial least squares

### Summary

Three-dimensional quantitative structure-activity relationship (3D QSAR) methods were applied on a series of inhibitors of HIV-1 integrase with respect to their inhibition of 3'-processing and 3'-end joining steps *in vitro*. The training set consisted of 27 compounds belonging to the class of thiazolothiazepines. The predictive ability of each model was evaluated using test set I consisting of four thiazolothiazepines and test set II comprised of seven compounds belonging to an entirely different structural class of coumarins. Maximum Common Substructure (MCS) based method was used to align the molecules and this was compared with other known methods of alignment. Two methods of 3D QSAR: comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) were analyzed in terms of their predictive abilities. CoMSIA produced significantly better results for all correlations. The results indicate a strong correlation between the inhibitory activity of these compounds and the steric and electrostatic fields around them. CoMSIA models with considerable internal as well as external predictive ability were obtained. A poor correlation obtained with hydrophobic field indicates that the binding of thiazolothiazepines to HIV-1 integrase is mainly enthalpic in nature. Further the most active compound of the series was docked into the active site using the crystal structure of integrase. The binding site was formed by the amino acid residues 64-67, 116, 148, 151-152, 155-156, and 159. The comparison of coefficient contour maps with the steric and electrostatic properties of the receptor shows high level of compatibility.

### Introduction

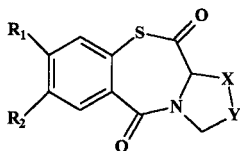
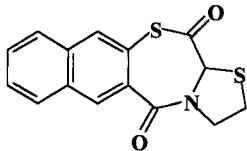
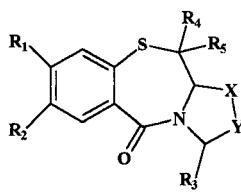
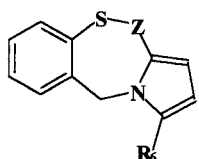
Several key enzymes involved in the replication cycle of human immunodeficiency virus (HIV) can be targeted for chemotherapeutic intervention, most notably, reverse transcriptase and protease [1]. Combination antiviral therapy with protease and reverse transcriptase inhibitors has demonstrated the potential therapeutic efficacy of antiviral therapy for treatment of acquired immunodeficiency syndrome (AIDS) [2]. However, the chronic nature of the infection and the high mutation rate of the virus have made it apparent that new antiretroviral agents are required to deal with the development and spread of resistance [3].

HIV-1 integrase (HIV-1 IN) is another such enzyme whose inhibition may be efficacious in the treatment of acquired immunodeficiency syndrome (AIDS), since this enzyme is required for viral replication, yet it is not indigenous to the human host [4–6].

HIV-1 integrase is an enzyme that mediates the integration of HIV-1 DNA into a host chromosome [7–9] and is essential to replication of the virus [10]. HIV-1 IN functions in a two-step manner by initially removing a dinucleotide unit from the 3'-ends of the viral DNA (termed 3'-processing). The 3'-processed strands are then transferred from the cytoplasm to the nucleus where they are introduced into the host DNA following 5-base pair offset cleavages of opposing host strands (termed 3'-strand transfer or end joining) [11]. This enzyme is therefore thought to be a suitable target

\*Author to whom correspondence should be addressed. E-mail: vithal@biogate.com

Table 1. Structures and activities of molecules in training set. (ref. 20).

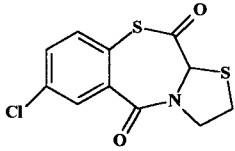
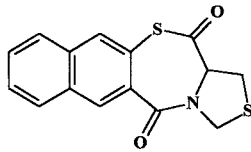
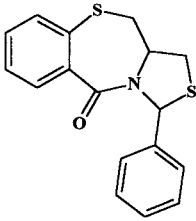
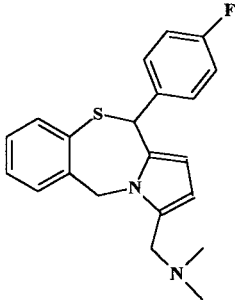
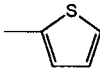
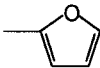
										
(1-17)		(18)		(19-24)		(25-27)				
Compd.	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	-X-Y-	Z	Integrase Assay IC <sub>50</sub> (mM)	
									3'-Processing	3'-End joining
1	H	H					-S-CH <sub>2</sub> -		0.110	0.146
2	H	Br					-S-CH <sub>2</sub> -		0.058	0.048
3	H	Me					-S-CH <sub>2</sub> -		0.064	0.055
4	H	H					-CH <sub>2</sub> -S-		0.208	0.227
5	H	Cl					-CH <sub>2</sub> -S-		0.134	0.135
6	H	Br					-CH <sub>2</sub> -S-		0.087	0.074
7	H	Me					-CH <sub>2</sub> -S-		0.052	0.063
8	NO <sub>2</sub>	H					-CH <sub>2</sub> -S-		0.090	0.100
9	H	OMe					-CH <sub>2</sub> -S-		0.215	0.200
10	OMe	OMe					-CH <sub>2</sub> -S-		0.650	0.331
11	H	H					-CH <sub>2</sub> -O-		0.333	0.333
12	OMe	OMe					-(CH <sub>2</sub> ) <sub>2</sub> -		1.000	1.000
13	H	H					-(CH <sub>2</sub> ) <sub>3</sub> -		0.450	0.343
14	H	Me					-S(O)-CH <sub>2</sub> -		0.590	0.590
15	H	H					-CH <sub>2</sub> -S(O)-		0.192	0.218
16	H	Cl					-CH <sub>2</sub> -S(O)-		0.237	0.240
17	H	OMe					-CH <sub>2</sub> -S(O)-		0.084	0.142
18									0.040	0.047
19	H	H	H	OAc	H		-(CH <sub>2</sub> ) <sub>2</sub> -		1.000	1.000
20	H	H	H	OMe	H		-(CH <sub>2</sub> ) <sub>2</sub> -		1.000	1.000
21	OMe	OMe	H	H	H		-(CH <sub>2</sub> ) <sub>2</sub> -		1.000	1.000
22	H	H	H	OH	H		-(CH <sub>2</sub> ) <sub>2</sub> -		1.000	1.000
23	H	H	H	OH	Ph		-(CH <sub>2</sub> ) <sub>2</sub> -		1.000	1.000
24	H	H	H	H	N((CH <sub>2</sub> ) <sub>2</sub> ) <sub>2</sub> NMe		-(CH <sub>2</sub> ) <sub>2</sub> -		1.000	1.000
25						H		>CO	0.590	0.300
26						H		>CHCO <sub>2</sub> Et	1.000	1.000
27						CH <sub>2</sub> N((CH <sub>2</sub> ) <sub>2</sub> ) <sub>2</sub> NMe		>CHPh	1.000	1.000

for chemotherapeutic intervention and has become a focus of anti-AIDS drug design efforts.

Although researchers have devoted much energy to inhibiting HIV-1IN, there are currently no integrase inhibitors in clinical development. Several classes of inhibitors have been reported to date, they are not without disadvantages. For example, a majority of hydroxylated aromatics such as catechol-containing compounds exhibit favorable inhibitory profiles in cell-free integrase assays, but have failed to exhibit antiviral activity due to dose-limiting toxicities in *in*

*vitro* assays [12]. Toxicities may reflect oxidation of the catechols to reactive quinone species, accounting for the observed losses in cell viability. Compounds containing this moiety can also cross-link protein, chelate metal, [13,14] and thus generally lack the desired selectivity [15]. The development of integrase inhibitors either lacking the catechol moiety or modifications which overcome its toxic properties have been the focus of considerable work [16]. Toward this goal, several non-catechol-containing structures like sulfonamides [17], diaryl sulfones [18], and aromatic

Table 2. Structures and activities of the molecules in test set I (ref. 20).

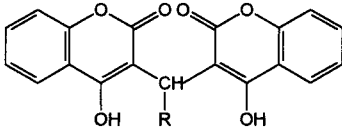
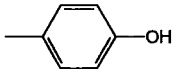
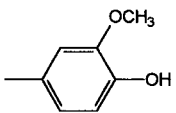
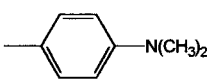
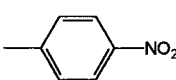
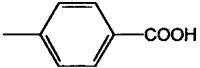
Compound		Integrase Assay	
		IC <sub>50</sub> (mM)	
		3'-Processing	3'-End joining
28		0.128	0.090
29		0.092	0.100
30		0.241	0.322
31		1.000	1.000
37		0.105	0.148
38		0.054	0.122

disulfides [19] have been identified, but very few of them exhibited antiviral activity. Thus, selectivity and in vivo potency remain difficult to achieve.

Recently, thiazolothiazepines were reported as selective integrase inhibitors, which possess antiviral activity [20]. To further explore the structural require-

ments of thiazolothiazepines as integrase inhibitors and their mechanism of action, two three-dimensional quantitative structure activity relationship (3D QSAR) methods: CoMFA (comparative molecular field analysis) and CoMSIA (comparative molecular similarity indices analysis) have been applied and compared with

Table 3. Structures and activities of coumarins used in test set II (from ref. 28).

			
Sr. No.	R	3'-Processing IC <sub>50</sub> (mM)	3'-Strand transfer IC <sub>50</sub> (mM)
32		0.134	0.074
33		0.187	0.092
34		0.094	0.069
35		0.054	0.018
36		0.057	0.051

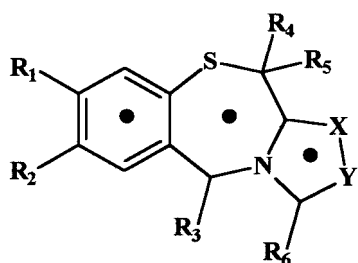


Figure 1. Centroid based alignment.

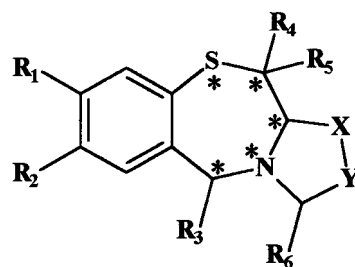


Figure 2. Atom based alignment.

respect to their predictive power. The CoMFA method of 3D QSAR was introduced by Cramer in 1988, in which an assumption is made that the interaction between an inhibitor and its molecular target is primarily noncovalent in nature and shape dependent. Therefore, a QSAR may be derived by sampling the steric and electrostatic fields surrounding a set of ligands and

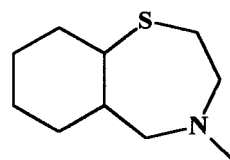


Figure 3. MCS based alignment.

Table 4. Summary of results from the CoMFA and CoMSIA analyses for 3'-processing.

	Alignment 1		Alignment 2		Alignment 3		Alignment 4	
	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA
$r_{cv}^2$	0.388	0.597	0.508	0.540	0.461	0.603	0.209	0.277
$S_{press}$	0.411	0.357	0.371	0.384	0.386	0.346	0.457	0.437
$r_{conv}^2$	0.825	0.945	0.844	0.913	0.832	0.913	0.694	0.630
SE	0.132	0.207	0.166	0.216	0.162	0.162	0.285	0.313
Comp.	3	6	3	6	3	5	2	2
F value	36.06	57.25	41.54	34.83	37.84	44.12	27.19	20.44
$Pr^2 = 0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Fraction:								
Steric	0.484	0.257	0.512	0.294	0.497	0.264	0.572	0.323
Electrostatic	0.516	0.743	0.488	0.706	0.503	0.736	0.428	0.677
$r_{pred}^2$ <sup>a</sup>	0.901	0.582	0.708	0.804	0.889	0.698	0.578	0.596
$r_{bs}^2$ <sup>b</sup>	0.897	0.962	0.909	0.952	0.901	0.934	0.791	0.720
SD <sup>b</sup>	0.039	0.019	0.034	0.025	0.037	0.029	0.059	0.068

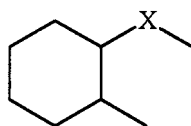
<sup>a</sup>Results using test set I.<sup>b</sup>Results from 1000 runs of bootstrapped analysis.

Figure 4. MCS for alignment of coumarins (X = O for coumarins; X = S for thiazolothiazepines).

correlating the differences in these fields to biological activity [21]. CoMFA calculates steric fields using a Lennard-Jones potential, and electrostatic fields using a Coulombic potential. While this approach has been widely accepted [22,23] and exceptionally valuable, it is not without problems. In particular, both potential functions are very steep near the van der Waals surface of the molecule, causing rapid changes in surface descriptions, and requiring the use of cut-off values so calculations are not done inside the molecular surface. In addition, a scaling factor is applied to the steric field, so that both fields can be used in the same PLS (partial least squares) analysis. Finally, changes in orientation of the superimposed molecule set, relative to the calculation grid, can cause significant changes in CoMFA results [24], again probably due to strict cut-off values.

The CoMSIA method of 3D QSAR was introduced by Klebe in 1994, in which using a common probe atom, similarity indices are calculated at regularly spaced grid points for the pre-aligned molecules. For the distance dependence between the probe atom and the molecule atoms a Gaussian function is used.

Because of the different shape of the Gaussian function, the similarity indices can be calculated at all grid points, both inside and outside the molecular surface. Although the major problem with both CoMFA and CoMSIA is the initial alignment of molecules, CoMSIA is not sensitive to changes in orientation of the superimposed molecules in the lattice, and the correlation results obtained by CoMSIA can be graphically interpreted in terms of field contribution maps allowing physicochemical properties relevant for binding to be easily mapped back onto molecular structures [25–27].

Although the CoMSIA approach presently considers five different property fields: steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor fields to focus on different physicochemical properties, the fields are highly interdependent on each other. Hence, to avoid duplication of fields and for direct comparison with CoMFA, only steric and electrostatic fields have been considered. Further, the hydrophobic field has been treated separately as well as along with steric and electrostatic fields to know its relevance to biological activity. Since hydrogen bond donor and acceptor properties are implicitly considered in electrostatic interactions, they have not been considered here.

Table 5. Results of analysis for 3'-processing with randomized biological activities and cross-validation using five groups.

	$r_{cv}^2$ <sup>a</sup>		$r_{cv}^2$ <sup>b</sup>	
	CoMFA <sup>c</sup>	CoMSIA <sup>d</sup>	CoMFA <sup>c</sup>	CoMSIA <sup>d</sup>
Mean	-0.12	-0.27	0.49	0.49
Std. Dev.	0.15	0.16	0.08	0.06
High	0.16	-0.03	0.60	0.58
Low	-0.54	-0.55	0.20	0.35

<sup>a</sup>Cross-validated  $r^2$  with randomized biological activity, average of 25 runs.

<sup>b</sup>Cross-validation using five groups with optimum number of components, average of 25 runs.

<sup>c</sup>Results using alignment 2.

<sup>d</sup>Results using alignment 3.

## Materials and methods

### Data set for Analysis

Published *in vitro* biological data for both 3'-processing and 3'-end joining (strand transfer) on a series of thiazolothiazepines were used for this study [20]. The structures and biological activities of the 27 molecules constituting the training set in the QSAR analyses are found in Table 1.

Test set I consists of 4 molecules (Table 2) and test set II consists of 7 molecules belonging to an entirely different structural class of coumarins (Table 3) [28]. Test set I molecules were selected to provide uniform sampling of biological activity in the class of thiazolothiazepines. Test set II was chosen to prove the versatility of the model. Although test set II compounds have some ionizable groups, they were modeled as neutral molecules for the purpose of this study.

The published  $IC_{50}$  values [20,28] whenever exact were taken as such. Whenever two values were given, the average of the two values was used to derive the model and in case of  $IC_{50}$  values with only lower limits, they were truncated to that particular value. This assumption will not affect the model derivation; since compounds with only lower limits (>1000) reported [20] are all considered to be inactive. This kind of assumption has been followed earlier [23,29].

All biological activities used in the present study were expressed as

$$pIC_{50} = -\log_{10} IC_{50}$$

where  $IC_{50}$  is the millimolar concentration of the inhibitor producing 50% inhibition.

## Computational Approaches

All molecular modeling techniques, CoMFA and CoMSIA studies described herein were performed on Silicon Graphics INDY R5000 workstation using the SYBYL 6.6 molecular modeling software from Tripos, Inc., St.Louis, MO [30].

### Molecular Conformation and Alignment

The compounds were built from fragments in the SYBYL database. Each structure was fully geometry optimized using the standard Tripos forcefield [31] with a distance dependent-dielectric function and a 0.001 kcal/mol · Å energy gradient convergence criterion. Partial atomic charges required for calculation of the electrostatic interaction were computed by a semiempirical molecular orbital method using the MOPAC 6.0 program. The charges were computed using the PM3 model Hamiltonian (keywords: 1SCF, RHF, MMOK) [32,33].

Using the systematic search protocol, rotatable bonds in compounds 24, 27 and 31 were searched from 0 to 360° in 10° increments. The lowest energy conformation of each compound (24, 27 and 31) thus obtained using systematic search, was minimized using Tripos forcefield and subsequently used in the analysis. Compound 18 was chosen as the template molecule because of its high potency and rigidity, on which other molecules were aligned.

### Alignment Rules

Superimposition of the molecules was carried out by the following four alignments.

1. Centroid-based alignment: In this, three ring centroids A, B and C were used for RMS-fitting to the template structure (Figure 1).
2. Atom-based alignment: In this, atoms 'SCCNC' of the molecules were used for RMS-fitting onto the corresponding atoms of the template structure. The atoms used for alignment are marked with asterisk (\*) in Figure 2.
3. Maximum Common Substructure-based alignment: Charisma in SYBYL 6.6 performs rigid alignment of molecules using maximum common substructure (MCS). It attempts to align molecules to a template molecule on a common backbone or core.

For each molecule to be aligned:

- a) The core is identified in the molecule.

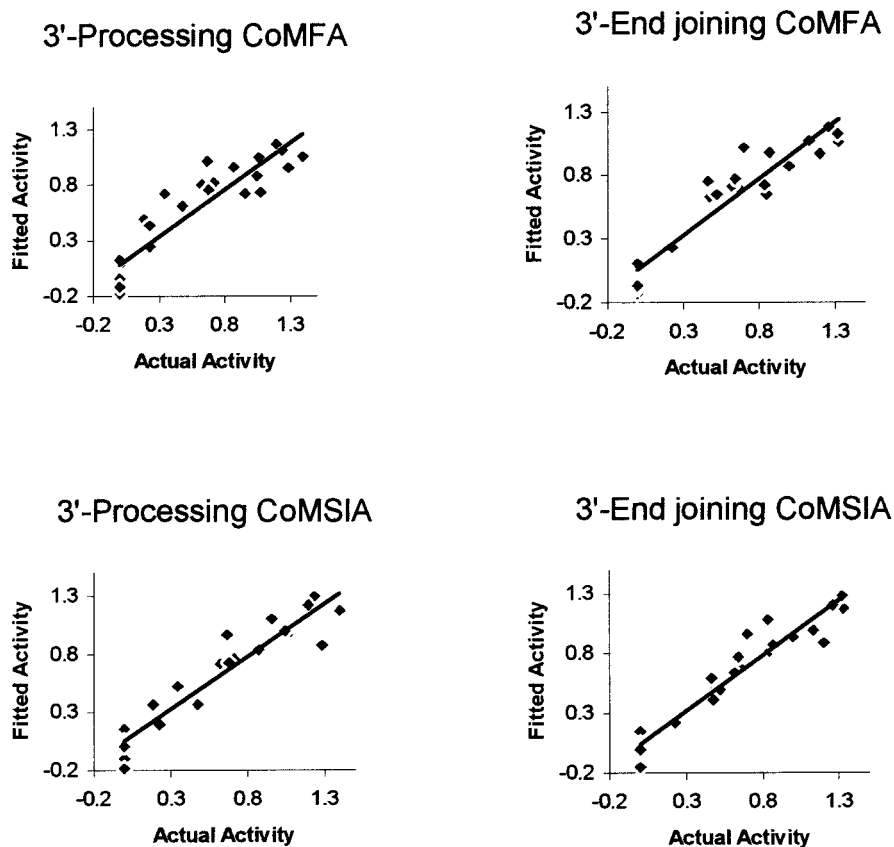


Figure 5. Fitted vs actual activity values for CoMFA and CoMSIA analysis of the training set. CoMFA results are from alignment 2 and CoMSIA from alignment 3.

Table 6. Summary of results from the CoMFA and CoMSIA analyses for 3'-end joining activity.

	Alignment 1		Alignment 2		Alignment 3		Alignment 4	
	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA
$r_{cv}^2$	0.575	0.740	0.675	0.718	0.594	0.744	0.308	0.406
$s_{press}$	0.328	0.266	0.290	0.287	0.320	0.267	0.410	0.380
$r_{conv}^2$	0.854	0.937	0.899	0.930	0.856	0.929	0.724	0.760
SE	0.192	0.132	0.160	0.143	0.191	0.140	0.259	0.246
comp.	3	5	3	6	3	5	2	3
F value	45.01	62.71	68.03	44.22	45.49	55.07	31.41	24.28
$Pr^2 = 0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Fraction:								
Steric	0.488	0.241	0.495	0.286	0.499	0.262	0.565	0.305
Electrostatic	0.512	0.759	0.505	0.714	0.501	0.738	0.435	0.695
$r_{pred}^2$ <sup>a</sup>	0.960	0.950	0.785	0.970	0.971	0.939	0.670	0.540
$r_{bs}^2$ <sup>b</sup>	0.905	0.946	0.934	0.953	0.908	0.939	0.804	0.849
SD <sup>b</sup>	0.035	0.021	0.026	0.022	0.036	0.025	0.061	0.048

<sup>a</sup>Results using test set I.

<sup>b</sup>Results from 1000 runs of bootstrapped analysis.

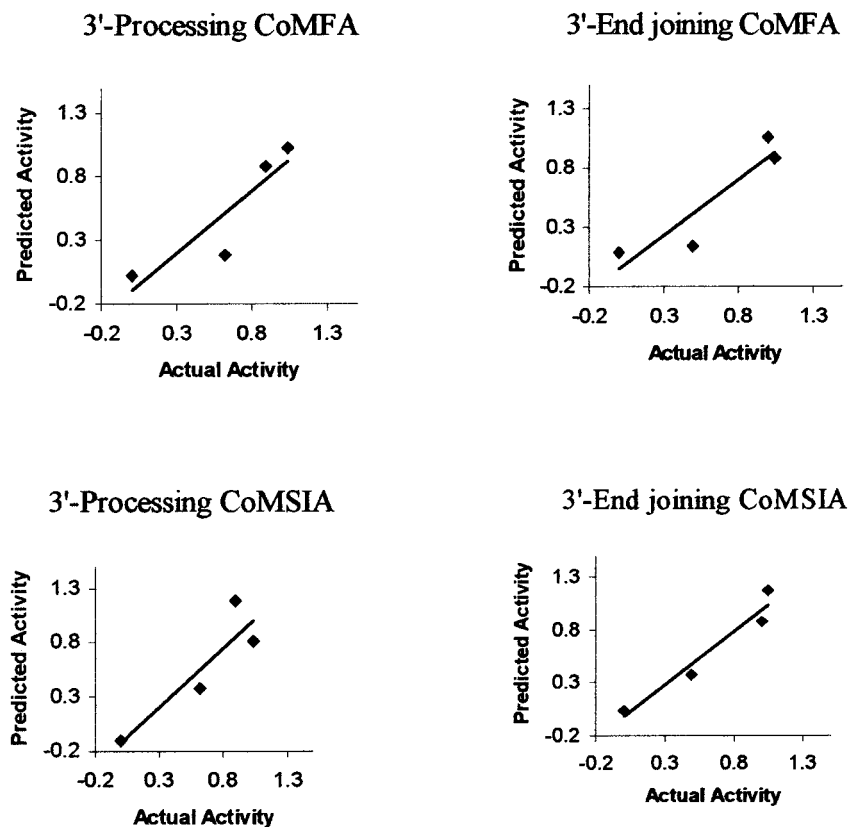


Figure 6. Predicted vs actual activity values for the CoMFA and CoMSIA analyses of test set I. CoMFA results are from alignment 2 and CoMSIA from alignment 3.

b) The core may be found more than once, or there may be more than one mapping of the core atoms to the molecule atoms. In this case a single mapping is chosen.

i) First the RMS deviation for each mapping is examined. The mapping with the lowest RMS deviation is chosen.

ii) If there is more than one mapping with the same RMS deviation (mappings are considered the same when the difference in RMS deviations is within 0.0005) the mapping that produces the lowest differential volume between the molecule and the template is chosen.

c) The molecule is fit to the template using the best mapping of the core to the molecule.

Using Charisma, by ignoring the bond types in the rings an MCS was determined which was subsequently used for alignment. MCS for training set and test set I is shown in Figure 3 and for test set II, it is shown in Figure 4.

4. Field fit alignment: This was carried out by using the SYBYL QSAR rigid body field fit command within SYBYL using compound 18 as a template. Field fit adjusts the alignment of the molecules such that their steric and electrostatic fields match the fields of the template molecule.

#### CoMFA and CoMSIA Analysis

For each of the alignment sets, CoMFA steric and electrostatic fields were separately calculated at each lattice intersection on a regularly spaced grid of 2.0 Å units in all x, y and z directions. The grid pattern, generated automatically by the SYBYL/CoMFA routine, extended 4.0 Å in all directions beyond the dimensions of the aligned molecules. The steric term represents the van der Waals (6-12) interactions, while the Coulombic term represents the electrostatic interactions for which a distance-dependent dielectric expression  $\epsilon = \epsilon_0 R_{ij}$  with  $\epsilon_0 = 1.0$  was adopted. An  $sp^3$  carbon atom with a van der Waals radius of 1.52 Å and a +1.0 charge was selected as the probe to calculate



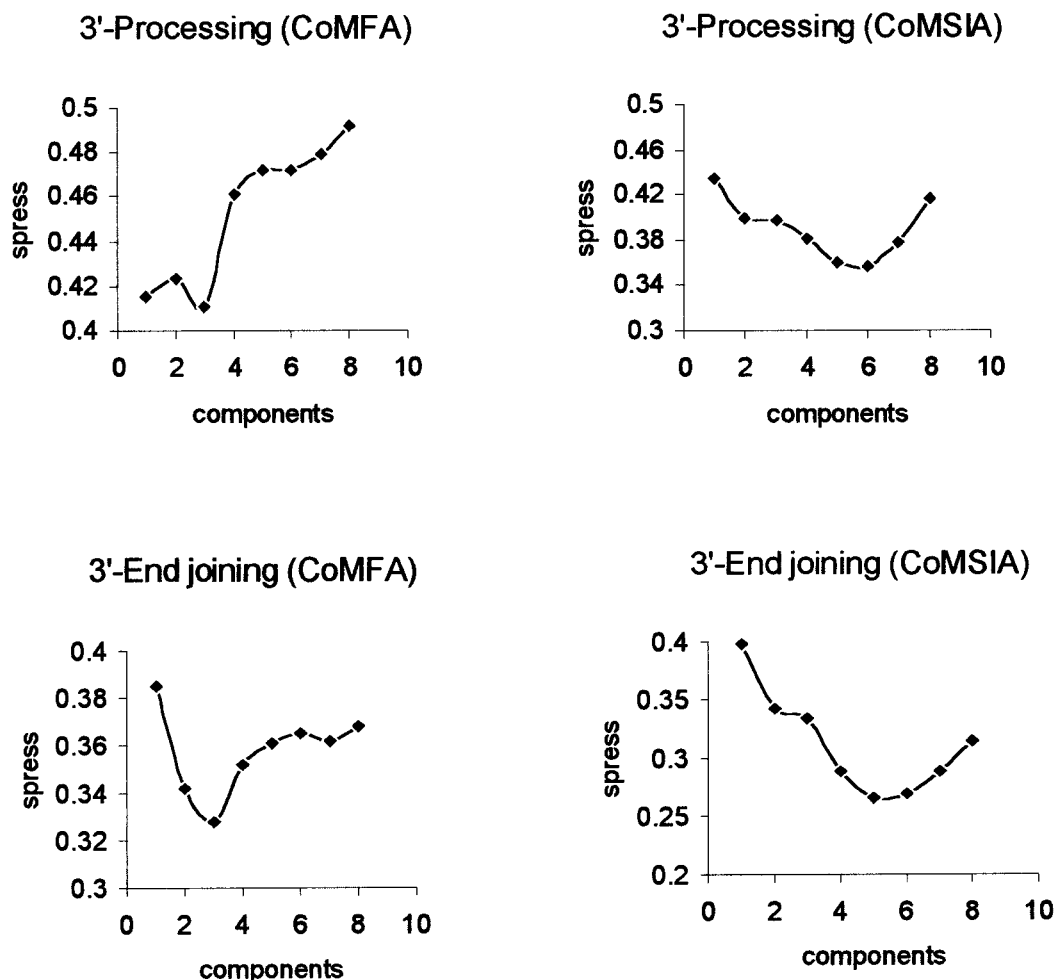


Figure 7. Plot of *spress* values versus components for alignment 1.

the steric and electrostatic fields. Values of the steric and electrostatic energy were truncated at 30 kcal/mol. The electrostatic contributions were ignored at lattice intersection with maximal steric interactions.

CoMSIA calculates similarity indices at the intersections of a surrounding lattice. The similarity index  $A_{F,k}$  [34] for a molecule *j* with *i* atoms at the grid point *q* is determined as follows:

$$A_{F,k}^q(j) = -\sum_i \omega_{\text{probe},k} \omega_{ik} e^{-\alpha r_{iq}^2}$$

where  $\omega_{ik}$  is the actual value of the physicochemical property *k* of atom *i*;  $\omega_{\text{probe},k}$  is the probe atom with charge +1, radius 1 Å and hydrophobicity = 1;  $r_{iq}$  is the mutual distance between probe atom at grid point *q* and atom *i* of the molecule. Three physicochemical properties *k* (steric, electrostatic and hydrophobic) were evaluated, using a common probe atom with

1 Å radius, charge and hydrophobicity. A Gaussian-type distance dependence was considered between the grid point *q* and each atom *i* of the molecule. The value of the so-called attenuation factor  $\alpha$  was set to 0.3. A lattice of 2 Å grid spacing was generated automatically.

#### Partial Least Squares (PLS) Analysis

To obtain a 3D QSAR, partial least squares (PLS) fitting was used [35]. The PLS method has been applied successfully in numerous QSAR studies aiming to rationalize those structural features affecting biological activity. PLS regression seeks a relationship between *Y* and *X*, where vector *Y* is the response or dependent variable and *X* represents the descriptor data.

PLS analyses were performed following the CoMFA standard implementation in SYBYL. The dif-

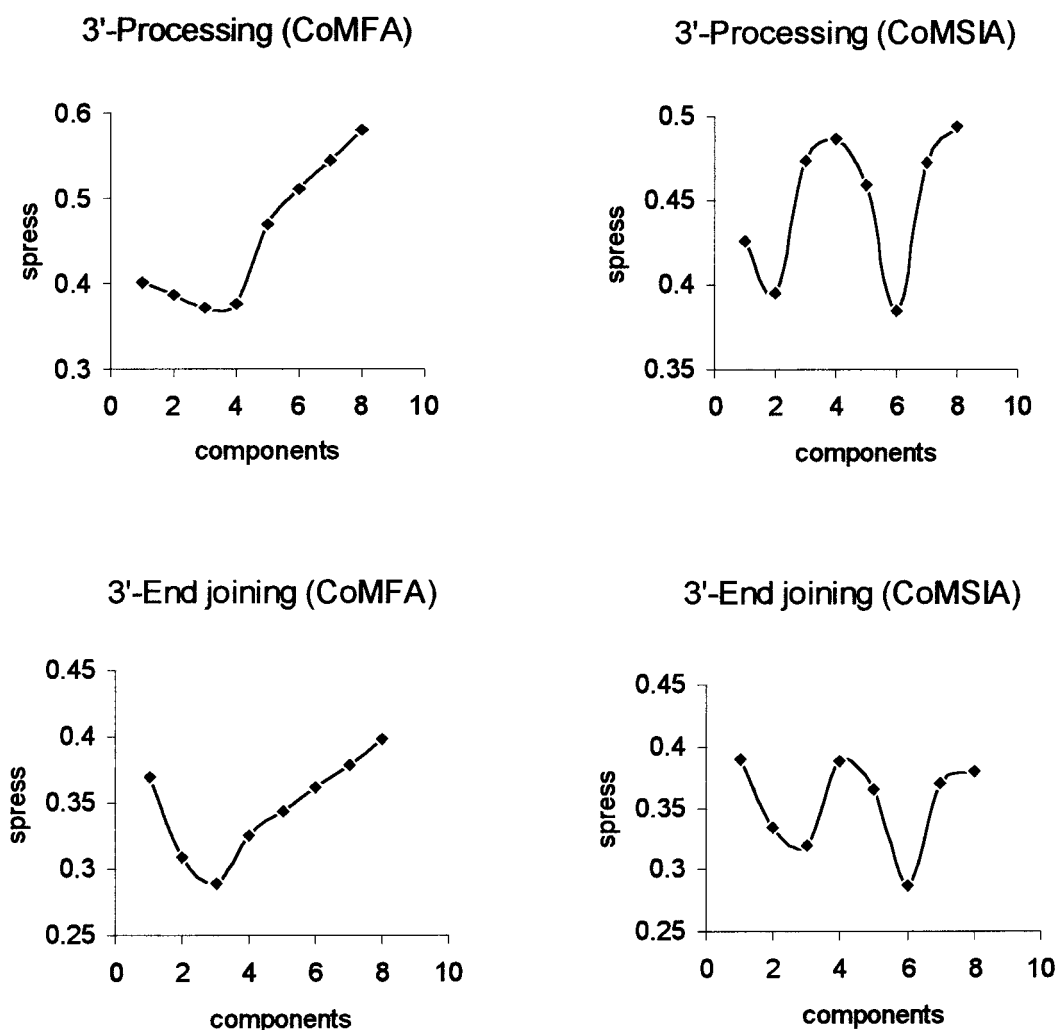


Figure 8. Plot of  $s_{press}$  values versus components for alignment 2.

ferent descriptor blocks have been scaled to each other using the CoMFA standard scaling option. To check statistical significance of the models, cross-validations were done by means of the 'leave-one-out' (LOO) procedure. The results from cross-validation analysis were expressed as the cross-validated  $r^2$  value ( $r_{cv}^2$ ). The cross-validated  $r^2$  is defined as

$$r_{cv}^2 = 1 - \text{PRESS} / \sum (Y - Y_{\text{mean}})^2$$

where  $\text{PRESS} = \sum (Y - Y_{\text{pred}})^2$ .

The optimal number of components was determined by selecting the smallest  $s_{press}$  value.  $s_{press}$  is the root mean PRedictive Error Sum of Squares. It is an expected uncertainty in prediction for an individ-

ual compound based on the data available from other compounds in the set.

$$s_{press} = (\text{PRESS} / (n - c - 1))^{1/2}$$

where  $n$  = number of rows and  $c$  = number of components. Usually the smallest  $s_{press}$  value corresponds to the highest  $r_{cv}^2$  value. The optimal number of components was subsequently used to derive the final QSAR models. For all conventional analyses (no cross-validation) the 'minimum sigma' standard deviation threshold was set to 2.0 kcal/mol. The  $r_{cv}^2$ ,  $s_{press}$ ,  $r_{conv}^2$  and SE values were computed as defined in SYBYL. SE is the standard error of estimate. It is a measure of the target property uncertainty still unexplained after the QSAR has been derived. In Tables 4 and 6,  $\text{Pr}^2 = 0$  means the probability of obtaining the

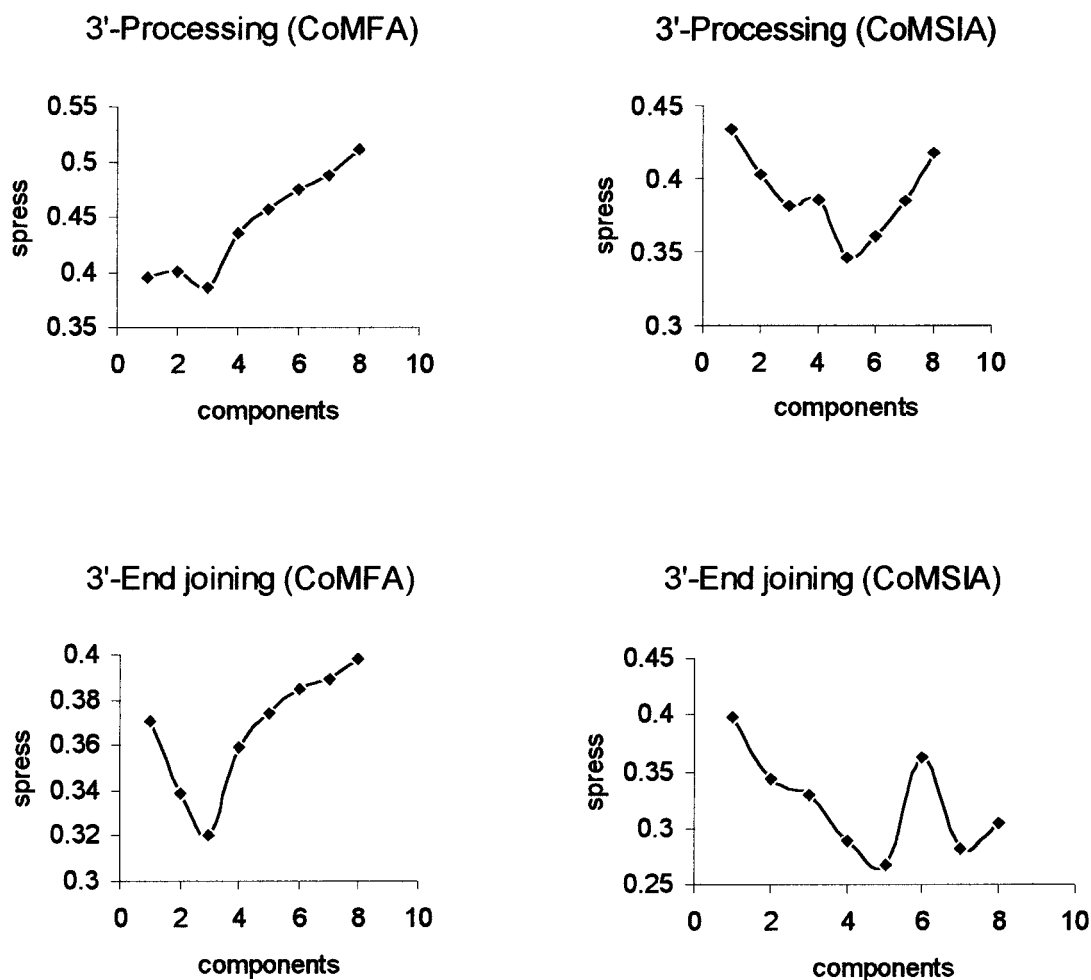


Figure 9. Plot of *spress* values versus components for alignment 3.

observed F-ratio value by chance alone, if the target and the explanatory variables themselves are truly uncorrelated. When  $Pr^2 = 0$  is zero, then results are not by chance and are significant. Additionally, to perform an even more rigorous statistical test, several runs of cross-validations using five groups were done in which each target property value is predicted by a model based on about 4/5, or 80% of the available data. To further assess the robustness and statistical confidence of the derived models, bootstrapping analysis (1000) runs was performed.

A common test to check the consistency of the models is to scramble the biological data and repeat the model derivation process, allowing detection of possible chance correlations. After randomizing our data set, we observed very low or negative  $r_{cv}^2$  values in the PLS analyses.

#### Predictive $r^2$ value

The predictive  $r^2$  was based only on molecules not included in the training set and is defined as:

$$r_{pred}^2 = (SD - PRESS)/SD$$

where SD is the sum of the squared deviations between the biological activity of molecules in the test set and the mean biological activity of the training set molecules and PRESS is the sum of the squared deviations between predicted and actual activity values for every molecule in the test set [36]. Like  $r_{cv}^2$ , the predictive  $r^2$  can assume a negative value reflecting a complete lack of predictive ability of the training set for the molecules included in the test set [37].

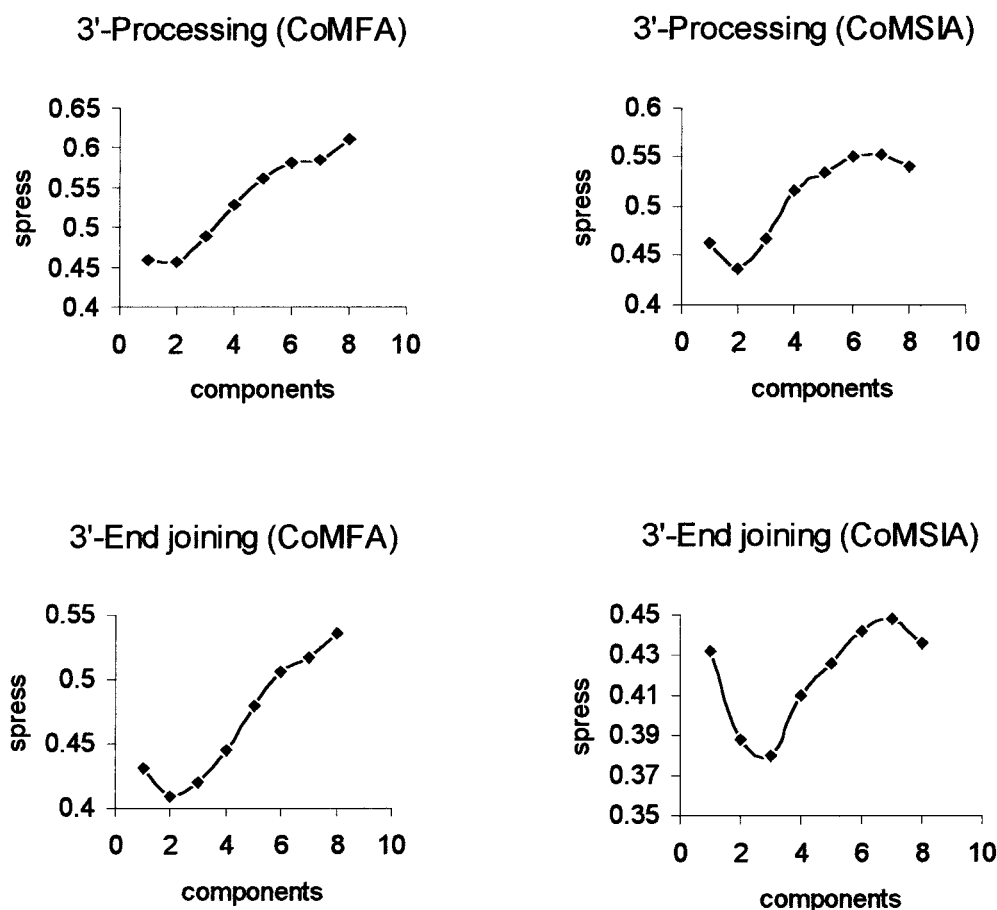


Figure 10. Plot of  $s_{press}$  values versus components for alignment 4.

#### Calculation of Additional Descriptors

Atom based partition coefficient (AlogP) values were calculated using Cerius2 version 3.5 [38] running on a Silicon Graphics O2 R5000 workstation.

## Results and Discussion

### 3'-Processing

The results of CoMFA and CoMSIA studies are summarized in Table 4. CoMFA analysis based on alignment 1 yielded a correlation with poor  $r_{cv}^2$  of 0.388 using three principal components. The conventional  $r^2$  for this analysis was 0.825. The model, although having poor internal predictivity, showed good external predictivity ( $r_{pred}^2 = 0.901$ ). CoMSIA studies using shape-based alignment yielded  $r_{cv}^2$  of 0.597 using 6 components and  $r_{conv}^2$  was 0.945. Although this

model yielded a satisfactory cross-validated  $r^2$ , the predictive ability for the test set molecules was poor ( $r_{pred}^2 = 0.582$ ).

CoMFA using alignment 2 produced an internally predictive ( $r_{cv}^2 = 0.508$  with 3 components) and statistically significant ( $r_{conv}^2 = 0.844$ ) model. The high bootstrapped  $r^2$  value ( $r_{bs}^2 = 0.909$ ) and small standard deviation suggest a high degree of confidence in the analysis. A good predictive ability with an  $r_{pred}^2 = 0.708$  for compounds in the test set was obtained. CoMSIA using atom-based alignment showed  $r_{cv}^2 = 0.540$  using 6 principal components and  $r_{conv}^2 = 0.913$ .

CoMFA using MCS-based alignment gave  $r_{cv}^2 = 0.461$  and  $r_{conv}^2 = 0.832$ . CoMSIA using this alignment yielded the model with the best internal predictivity ( $r_{cv}^2 = 0.603$ ) and statistical significance ( $r_{conv}^2 = 0.913$ ) with good external predictivity ( $r_{pred}^2 = 0.698$ ). To obtain statistical confidence limits, the non-cross-validated analysis was repeated with

1000 bootstrap groups, which yielded an  $r^2$  of 0.934 and standard deviation of 0.02. To further check the predictive ability of the model, a set of seven compounds belonging to an entirely different structural class: coumarins was predicted. For test set II (Table 3), the model showed  $r_{\text{pred}}^2 = 0.474$ . Thus, this model is versatile as it is able to predict the activities of molecules of a different class. MCS for the superimposition of coumarins on compound 18 is shown in Figure 4. The results of cross-validations using five groups and randomized 3'-processing activities are summarized in Table 5.

It was somewhat anticipated that, by aligning the molecules of the training set using the field-fit minimization procedure, the internal consistency of the model represented by the cross-validated  $r^2$  value would improve. This was not to be the case as decrease in  $r_{\text{cv}}^2$  value was noted with both CoMFA and CoMSIA.

### 3'-End joining

The results of CoMFA and CoMSIA studies are summarized in Table 6. The CoMFA model with shape-based alignment gave  $r_{\text{cv}}^2$  of 0.575 and  $r_{\text{conv}}^2$  of 0.854. CoMSIA using this alignment showed  $r_{\text{cv}}^2 = 0.740$ ,  $r_{\text{conv}}^2 = 0.937$ , and  $r_{\text{pred}}^2 = 0.950$ .

Using alignment 2, CoMFA yielded a model with good internal and external predictivity ( $r_{\text{cv}}^2 = 0.675$ ,  $r_{\text{conv}}^2 = 0.899$ , and  $r_{\text{pred}}^2 = 0.785$ ). Steric and electrostatic contributions for this model were 0.495 and 0.505 respectively. High F-value and  $r_{\text{bs}}^2$  along with low  $s_{\text{press}}$  value place further confidence in this model. CoMSIA with this alignment yielded model with  $r_{\text{cv}}^2 = 0.718$  and  $r_{\text{conv}}^2 = 0.930$ .

CoMFA and CoMSIA models using the MCS-based alignment yielded similar results to those of shape-based alignment. Finally, the external predictivity of this model was checked on test set II which showed  $r_{\text{pred}}^2 = 0.518$ . The results of cross-validations using five groups and randomized biological activities with this model are summarized in Table 7.

Like 3'-processing, poor results were obtained for 3'-end joining with field-fit alignment procedure.

The observed and calculated  $\text{pIC}_{50}$  values for training set, test set I, and test set II are given in Table 8, Table 9, and Table 10 respectively. In all the above cases, the number of PLS-components is higher in CoMSIA than CoMFA. Probably this observation results from the significantly higher number of lattice points showing steadily varying field values (e.g., inside the

Table 7. Results of analysis for 3'-end joining with randomized biological activities and cross-validation using five groups.

	$r_{\text{cv}}^2$ <sup>a</sup>		$r_{\text{cv}}^2$ <sup>b</sup>	
	CoMFA <sup>c</sup>	CoMSIA <sup>d</sup>	CoMFA <sup>c</sup>	CoMSIA <sup>d</sup>
Mean	-0.23	-0.29	0.63	0.65
Std. Dev.	0.22	0.21	0.09	0.08
High	0.20	-0.02	0.73	0.75
Low	-0.74	-0.83	0.36	0.43

<sup>a</sup>Cross-validated  $r^2$  with randomized biological activity, average of 25 runs.

<sup>b</sup>Cross-validation using five groups with optimum number of components, average of 25 runs.

<sup>c</sup>Results using alignment 2.

<sup>d</sup>Results using alignment 3.

molecules). The optimal number of components was selected on the basis of lowest  $s_{\text{press}}$  values. The plots of fitted versus actual activity values for the training set molecules and predicted versus actual activity values for the test set molecules are shown in Figure 5 and Figure 6 respectively. The plots of  $s_{\text{press}}$  versus number of components for four alignments are given in Figures 7, 8, 9, and 10 respectively.

A previous CoMFA study on HIV-1 integrase inhibitors belonging to the class of flavones performed by Raghavan et al, shows the importance of steric (21% and 20.5% for 3'-processing and 3'-end joining respectively) and electrostatic (79% and 79.5% for 3'-processing and 3'-end joining respectively) fields for anti-HIV-1 IN activity [39]. Our model using MCS based alignment with CoMSIA shows steric contribution of 26.4% and 26.2% for 3'-processing and 3'-end joining; and electrostatic contribution of 73.6% and 73.8% for 3'-processing and 3'-end joining respectively (Tables 4 and 6). Thus, both models show almost comparable results indicating a major influence of electrostatic interactions on HIV-1 IN inhibitory activity of these compounds.

### Role of Hydrophobic Interactions

Structure-activity relationship studies of thiazolothiazepines as reported by Neamati et al. [20], revealed that compounds with a naphthalene ring system possess enhanced potency, suggesting that a hydrophobic pocket in the integrase active site might accommodate an aromatic system. So, in order to gain further insights into the role of hydrophobic interactions, hydrophobic fields were computed and correlated alone as well as in combination with the steric and elec-

Table 8. Observed versus calculated pIC<sub>50</sub> values for molecules in the training set.

Compd.	3'-Processing			3'-End joining		
	obsd. pIC <sub>50</sub>	calcd. pIC <sub>50</sub>		obsd. pIC <sub>50</sub>	calcd. pIC <sub>50</sub>	
		CoMFA <sup>a</sup>	CoMSIA <sup>b</sup>		CoMFA <sup>a</sup>	CoMSIA <sup>b</sup>
1	0.958	0.722	1.103	0.835	0.722	1.081
2	1.236	1.110	1.295	1.318	1.129	1.282
3	1.193	1.166	1.220	1.259	1.179	1.200
4	0.681	0.755	0.730	0.643	0.772	0.773
5	0.871	0.958	0.835	0.869	0.979	0.871
6	1.060	1.050	0.964	1.130	1.071	0.989
7	1.283	0.955	0.874	1.200	0.969	0.887
8	1.045	0.879	1.000	1.000	0.871	0.933
9	0.667	1.017	0.965	0.698	1.019	0.958
10	0.187	0.495	0.366	0.479	0.612	0.534
11	0.477	0.611	0.364	0.477	0.631	0.410
12	0.000	0.077	0.084	0.000	0.122	0.120
13	0.346	0.721	0.524	0.464	0.754	0.593
14	0.229	0.245	0.227	0.229	0.233	0.218
15	0.715	0.819	0.763	0.660	0.708	0.675
16	0.624	0.800	0.720	0.619	0.715	0.642
17	1.073	0.735	0.945	0.847	0.651	0.819
18	1.397	1.058	1.173	1.327	1.061	1.177
19	0.000	-0.12	0.071	0.000	-0.10	0.061
20	0.000	-0.07	-0.12	0.000	-0.05	-0.10
21	0.000	-0.17	-0.09	0.000	-0.13	-0.11
22	0.000	0.083	0.157	0.000	0.096	0.148
23	0.000	-0.02	-0.18	0.000	-0.02	-0.15
24	0.000	-0.04	0.051	0.000	-0.08	0.050
25	0.229	0.437	0.195	0.522	0.649	0.494
26	0.000	-0.11	0.042	0.000	-0.06	0.036
27	0.000	0.126	0.004	0.000	0.110	-0.005

<sup>a</sup>Results using alignment 2.<sup>b</sup>Results using alignment 3.Table 9. Observed versus predicted pIC<sub>50</sub> values for molecules in the test set I.

Compd.	3'-Processing			3'-End joining		
	obsd. pIC <sub>50</sub>	calcd. pIC <sub>50</sub>		obsd. pIC <sub>50</sub>	calcd. pIC <sub>50</sub>	
		CoMFA <sup>a</sup>	CoMSIA <sup>b</sup>		CoMFA <sup>a</sup>	CoMSIA <sup>b</sup>
28	0.892	0.882	1.195	1.045	0.889	1.178
29	1.036	1.030	0.817	1.000	1.065	0.880
30	0.617	0.182	0.383	0.492	0.145	0.377
31	0.000	0.025	-0.10	0.000	0.086	0.035

<sup>a</sup>Results using alignment 2.<sup>b</sup>Results using alignment 3.

trostatic fields to the 3'-processing and 3'-end joining activities. The results of this analysis are summarized in Table 11. With hydrophobic fields alone, poor  $r_{cv}^2$  values were obtained, and along with steric and electrostatic fields  $r_{cv}^2$  decreased as compared to when only

steric and electrostatic fields were used. To further consolidate this view, AlogP values were correlated with 3'-processing and end joining, alone as well as in combination with the steric and electrostatic fields, but the results were no better than those obtained with hy-

Table 10. Observed versus predicted pIC<sub>50</sub> values for molecules in the test set II.

Compd.	3'-Processing			3'-End joining		
	obsd. pIC <sub>50</sub>	calcd. pIC <sub>50</sub>		obsd. pIC <sub>50</sub>	calcd. pIC <sub>50</sub>	
		CoMFA <sup>a</sup>	CoMSIA <sup>b</sup>		CoMFA <sup>a</sup>	CoMSIA <sup>b</sup>
32	0.87	0.020	0.217	1.13	0.056	0.316
33	0.73	0.056	0.805	1.04	0.106	0.800
34	1.03	0.013	0.247	1.16	0.039	0.353
35	1.27	0.033	0.262	1.74	0.100	0.374
36	1.24	0.035	0.258	1.29	0.104	0.371
37	0.98	0.172	0.280	0.83	0.223	0.401
38	1.26	0.320	0.422	0.91	0.305	0.450

<sup>a</sup>Results using alignment 2.<sup>b</sup>Results using alignment 3.

Table 11. Results of analysis with Hydrophobic field.

	3'-Processing <sup>a</sup>		3'-End joining <sup>a</sup>	
	Hydrophobic	CoMSIA <sup>b</sup>	Hydrophobic	CoMSIA <sup>b</sup>
r <sub>cv</sub> <sup>2</sup>	0.157	0.286	0.179	0.453
s <sub>press</sub>	0.463	0.426	0.437	0.364
r <sub>conv.</sub> <sup>2</sup>	0.574	0.519	0.578	0.835
comp.	1	1	1	2

<sup>a</sup>CoMSIA analysis using alignment 3..<sup>b</sup>Results using steric, electrostatic and hydrophobic fields together.

drophobic fields. The results of analysis using AlogP values are summarized in Table 12. The only end point that has been found useful in deciding the validity of any CoMFA or CoMSIA model is a high enough r<sub>cv</sub><sup>2</sup>. With hydrophobic field and AlogP, poor r<sub>cv</sub><sup>2</sup> values were obtained. The 95% confidence limit for r<sub>cv</sub><sup>2</sup> in CoMFA or CoMSIA is about 0.3, hence the role of hydrophobic interactions is not statistically significant as r<sub>cv</sub><sup>2</sup> values less than 0.3 were observed [40]. Thus, an additional aromatic ring may not be participating in hydrophobic interactions but may be involved in a cation- $\pi$  type of interaction with a cationic residue in the active site of an enzyme [41]. This type of stabilizing interaction between a cation and the electrons of the aromatic ring system has been observed for aromatic residues in other protein structures [42] and may explain the enhanced activities of compounds, which possess the naphthalene ring system. Poor r<sub>cv</sub><sup>2</sup> values obtained with hydrophobic fields also indicate that the areas that get desolvated during binding are likely not hydrophobic patches but are rather polar.

Table 12. Results of analysis with AlogP values.

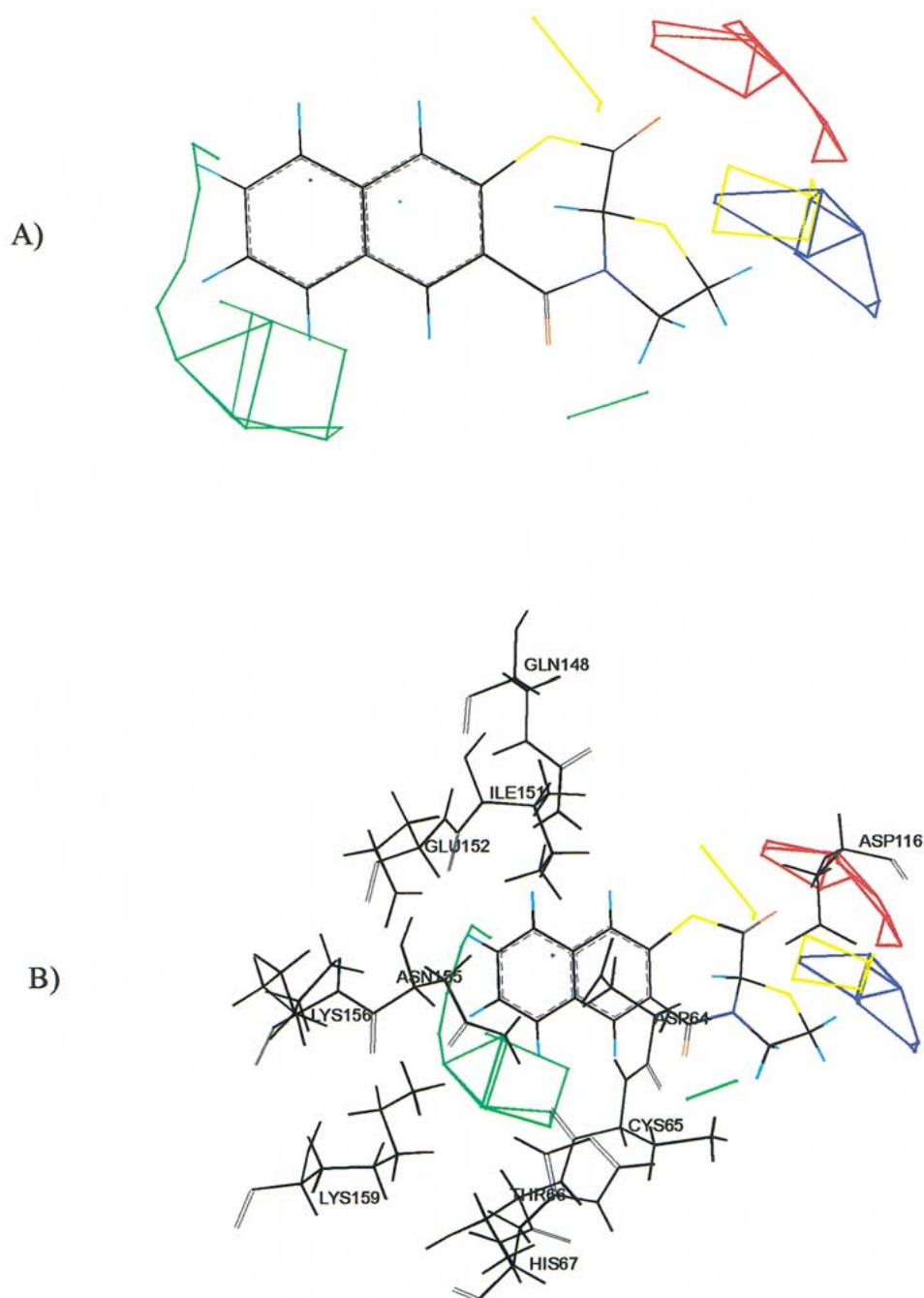
	3'-Processing <sup>a</sup>		3'-End joining <sup>a</sup>	
	AlogP	CoMSIA <sup>b</sup>	AlogP	CoMSIA <sup>b</sup>
r <sub>cv</sub> <sup>2</sup>	-0.107	-1.84e <sup>04</sup>	-0.133	0.563
s <sub>press</sub>	0.530	69.84	0.513	0.366
r <sub>conv.</sub> <sup>2</sup>	0.027	0.692	0.014	0.974
comp.	1	2	1	7

<sup>a</sup>CoMSIA analysis using alignment 3..<sup>b</sup>Results using steric, electrostatic and AlogP values together.

### Graphical Interpretation of the Results

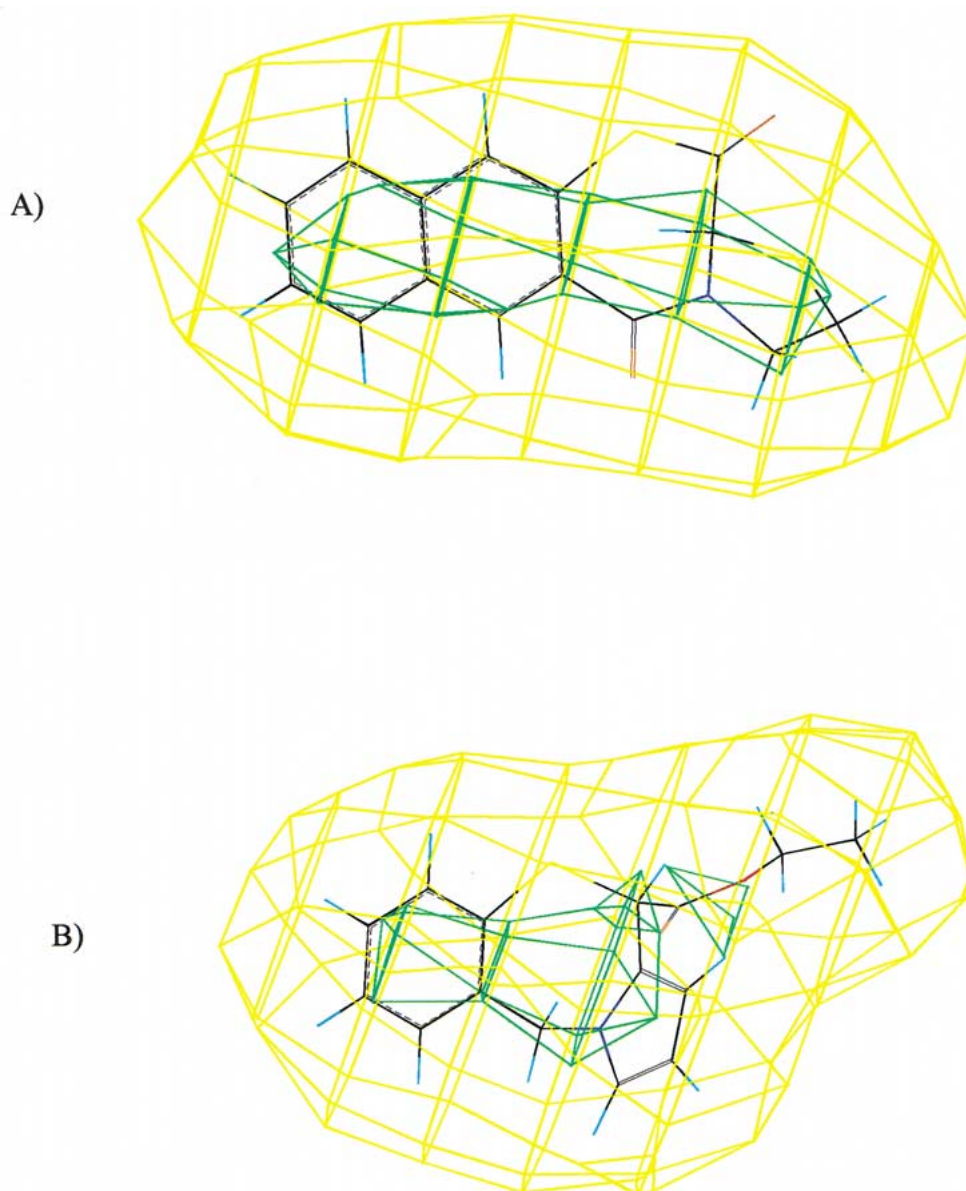
In order to visualize the information content of the derived 3D-QSAR model, CoMFA contour maps were generated by interpolating the products between the 3D-QSAR coefficients and their associated standard deviations. Figure 11a shows the CoMFA contour map in the form of mesh from the analysis based on alignment 2 using compound 18 as a reference structure. The green and yellow polyhedra describe regions of space around the molecules where increases in steric bulk respectively enhance or diminish inhibition of integrase. The red and blue polyhedra describe regions where a high electron density (i.e., negative charge or polarity) within the substrate structure enhances or diminishes activity respectively.

Besides greater robustness and better predictive power, the CoMSIA method provides significantly improved contour diagrams. They allow the correlation results to be mapped back onto molecular structures. The coefficient contour maps using the field type 'STDEV\*COEFF' were generated. Since both 3'-processing and end joining are sequential in enzymatic catalysis and involve similar reaction mecha-



*Figure 11.* A) The CoMFA stdev\*coeff contour plot from the analysis based on alignment 2. Sterically favored areas (contribution level of 80%) are represented by green polyhedra. Sterically disfavored areas (contribution level of 20%) are represented by yellow polyhedra. Negative charge favored areas (contribution level of 80%) are represented by blue polyhedra. Negative charge disfavored areas (contribution level of 20%) are represented by red polyhedra. B) Superimposition of CoMFA contour plots on active site of HIV-1 IN are depicted.

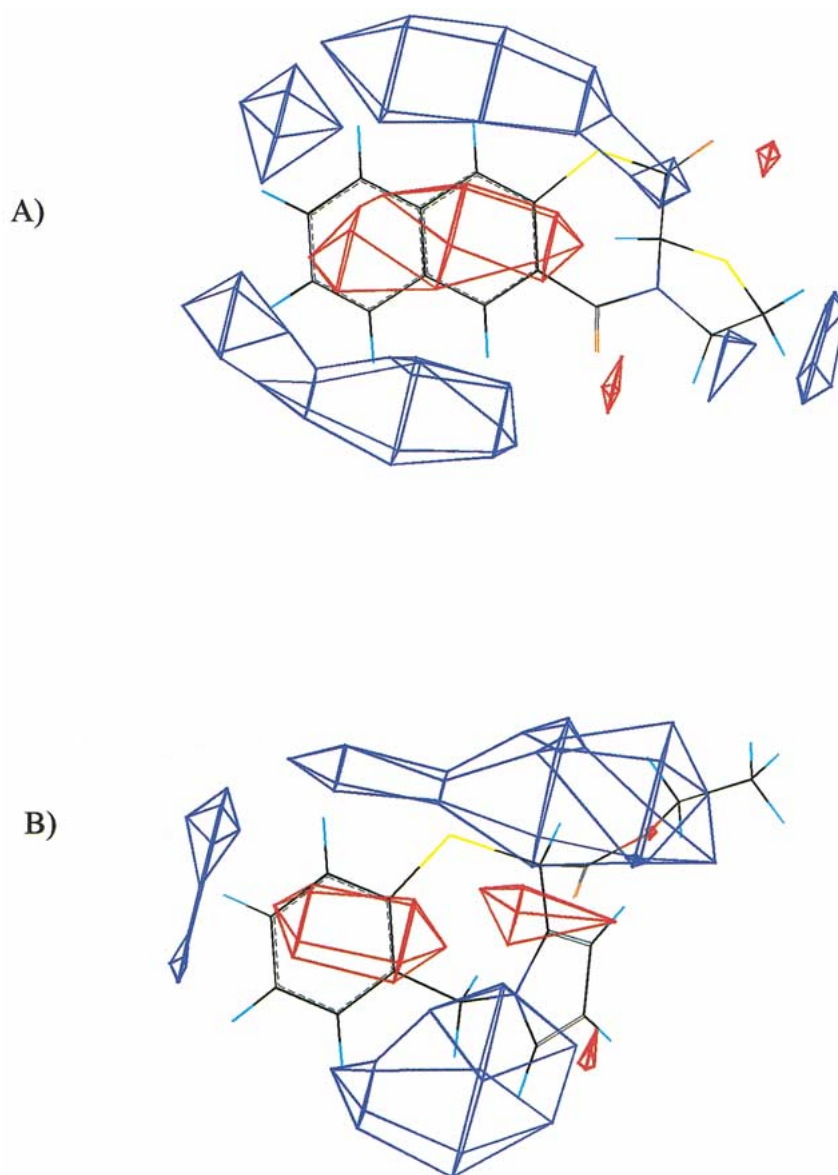




*Figure 12.* Steric maps from the CoMSIA model using alignment 3 for inhibition of 3'-processing. (A) Compound 18 shown inside the field. (B) Compound 26 shown inside the field. Green contours (80% contribution) enclose areas where steric bulk will enhance affinity and yellow contours (20% contribution) highlight areas, which should be kept unoccupied; otherwise binding affinity will decrease.

nism, contour maps of both are same and only those of 3'-processing are shown here. The steric contribution contour plots from the analysis based on alignment 3 for 3'-processing are plotted in Figure 12. In this figure, the green contours represent regions of high steric tolerance (80% contribution), while the yellow contours represent regions of low steric bulk tolerance (20% contribution). From a small sterically favorable region and a very large sterically disfavored region

surrounding the molecule, it appears that the size of the binding site in the enzyme is limited. The green region in the interior of molecule arises from the fact that in CoMSIA fields are calculated inside as well as outside the molecular surface. But, the regions of molecules inside its van der Waals radius are not involved in drug – receptor interactions. Certain compounds project their bulky steric groups in this sterically disfavored region and hence are inactive. For



*Figure 13.* Electrostatic maps from the COMSIA model using alignment 3 for inhibition of 3'-processing. (A) Compound 18 shown inside the field. (B) Compound 26 shown inside the field. Blue contours (80% contribution) encompass regions where an increase of positive charge will enhance affinity, whereas in red contoured areas (20% contribution) more negative charges are favorable for binding properties.

example, compound 26 projects its ethyl group totally in this yellow region and hence, its inactivity can be rationalized on the basis of its steric contour plot shown in Figure 12b.

The electrostatic contour plots for 3'-processing are shown in Figure 13. In this figure, the red contours represent regions of decreased tolerance for positive charge (20% contribution), while the blue contours represent regions of increased tolerance for positive

charge (80% contribution). The presence of a red contour in the vicinity of the amide carbonyl oxygen corroborates its role in a hydrogen bonding interaction. The activity of certain compounds, which possess Cl, Br or NO<sub>2</sub> group on the aromatic ring, can be rationalized due to the presence of blue contour surrounding that ring. Further, the presence of small red contour in the terminal aromatic ring of compound 18 supports the hypothesis of its involvement in cation- $\pi$  type of

interaction with the enzyme. The steric and electrostatic contour plots of compound 18 and compound 26 for 3'-processing are shown for reference.

#### *Comparison of Contour Plots with Active Site of HIV-1 Integrase*

The three-dimensional structure of HIV-1 IN complexed with thiazolothiazepines has not been determined. Hence, the structure of HIV-1 IN catalytic domain complexed with a different inhibitor 1-(5-chloroindol-3-yl)-3-hydroxy-3-(2H-tetrazol-5-yl)-pro-

penone (5CITEP) (1QS4.pdb) was used [43]. Only subunit A was used in the present study, since it was found to be complexed with the inhibitor 5CITEP. To this subunit complexed with 5CITEP, all the hydrogens were added and was subjected to a refinement protocol in which the constraints on the enzyme were gradually removed and minimized until the rms gradient was 0.05 kcal/mol Å [44,45]. The energy minimization was carried out using Tripos force field within SYBYL 6.6. This energy minimized structure without any metal ion was used for docking studies. To find a sterically reasonable binding geometry for specific interactions of thiazolothiazepines in the active site of HIV-1 IN, the docking option available within SYBYL 6.6 with default settings was used. Active site was defined using a radius of 12 Å around 5CITEP. Employing interactive docking procedure, compound 18 was docked into the enzyme active site, the initial placement of which was guided by the crystallographic position of 5CITEP. After docking, this complex was subjected to molecular dynamics (MD) simulations at 300°K using NVT ensemble with equilibration for 10ps, followed by simulation for 50ps with a time step of 1fs. Trajectory frames were saved after every 50fs. A distance dependent dielectric of 4.00 was used throughout the calculations [46]. The lowest potential energy conformer from MD simulations was further minimized using Powell method to a gradient of 0.05 kcal/mol Å. This cycle of interactive docking, MD simulations, and minimization was repeated for compound 18 till no further lowering of interaction energy (−47.394 kcal/mol) was observed in docking. The docked compound 18 was found to form two H-bonds, one with backbone −NH− of CYS65 (2.278 Å) and other with backbone −NH− of ASP116 (2.72 Å).

The contour maps are  $\frac{3}{4}$  consistent with the binding site of HIV-1 IN. The comparison of contour plots

Table 13. Overlap of the CoMFA contour maps and HIV-1 integrase active site residues.

	CoMFA regions	Active site residues
1	Sterically favored (green)	ASP64, CYS65, THR66, GLU152, ASN155, LYS159
2	Sterically forbidden (yellow)	ASP116
3	Positive charge favoring (blue)	ASP64, ASP116
4	Negative charge favoring (red)	ASP116

and the active site of integrase should be viewed very carefully as contour plots should not be overinterpreted as receptor maps [21]. The superimposition of CoMFA contour plots on active site residues is shown in Figure 11b. The overlapping residues of the integrase active site are given in Table 13. The sterically favorable green region occupies the clefts in the enzyme, whereas sterically unfavorable region overlaps with atoms of ASP116. Blue contours were observed near ASP64 and ASP116 that are favorable to the positive charge, and red contours were observed near ASP116 that favor negative charge.

#### **Conclusions**

Two 3D QSAR methods i.e., CoMFA and CoMSIA have been applied to a set of recently described thiazolothiazepine inhibitors of HIV-1 integrase. The present study has shown CoMSIA to possess better predictive power and greater robustness compared to CoMFA. The resulting 3D QSAR models show good correlations between steric and electrostatic field and HIV-1 integrase inhibitory activities. The contour diagrams obtained for the CoMSIA field contributions can be mapped back onto structural features accounting for the activity trends among the inhibitors. The 3D QSAR culminating from the training set yielded a regression equation with a high degree of statistical significance and performed well in predicting the biological activities of compounds in the test set. On the basis of the spatial arrangement of the field contributions, novel molecules can be designed that are predicted to possess improved biological activity. 3D QSAR models are a direct mirror of the structural variations inherently present in the selected data set. Accordingly, the selection of a structurally more diverse set of inhibitors should allow mapping features distinct from those highlighted by the present study.

## Acknowledgements

The authors gratefully acknowledge support for this research from the University Grants Commission (UGC), New Delhi, under its DSA and COSIST programmes. MM thanks UGC for the award of senior research fellowship.

## References

1. Hariprasad, V., Talele, T.T., Kulkarni, V.M., *Pharm. Pharmacol. Commun.*, 4 (1998) 365.
2. De Clercq, E., *J. Med. Chem.*, 38 (1995) 2491.
3. Richman, D.D., *Nature*, 410 (2001) 995.
4. Sakai, H., Kawamura, M., Sakuragi, J., Sakuragi, S., Shibata, R., Ishimoto, A., Ono, H., Ueda, S., Adachi, A., *J. Virol.*, 67 (1993), 1169.
5. Taddeo, B., Haseltine, W. A., Farnet, C.M., *J. Virol.*, 68 (1994) 8401.
6. Engelman, A., Englund, G., Orenstein, J.M., Martin, M.A., Craigie, R., *J. Virol.*, 69 (1995) 2729.
7. Goff, S.P., *Annu. Rev. Genet.*, 26 (1992) 527.
8. Vink, C., Plasterk, R.H.A., *Trends Genet.*, 9 (1993) 433.
9. Craigie, R., *Trends Genet.*, 8 (1992) 187.
10. LaFemina, R.L., Schneider, C.L., Robibins, H.L., Callahan, P.L., LeGrow, K., Roth, E., Schleif, W.A., Emini, E.A., *J. Virol.*, 66 (1992) 7414.
11. Craigie, R., *Trends Genet.*, 8 (1992) 187.
12. Stanwell, C., Ye, B., Yuspa, S.H., Burke, T.R., Jr., *Biochem. Pharmacol.*, 52 (1996) 475.
13. Neamati, N., Hong, H., Owen, J.M., Sunder, S., Winslow, H.E., Christensen, J.L., Zhao, H., Burke, J.T.R., Milne, G.W.A., Pommier, Y., *J. Med. Chem.*, 41 (1998) 3202.
14. Neamati, N., Hong, H., Mazumder, A., Wang, S., Sunder, S., Nicklaus, M. C., Milne, G.W., Proksa, B., Pommier, Y., *J. Med. Chem.*, 40 (1997) 942.
15. Pommier, Y., Pilon, A., Bajaj, K., Mazumder, A., Neamati, N., *Antiviral Chem. Chemother.*, 8 (1997) 483.
16. Neamati, N., Sunder, S., Pommier, Y., *Drug Discovery Today*, 2 (1997) 487.
17. Nicklaus, M.C., Neamati, N., Hong, H., Mazumder, A., Sunder, S., Chen, J., Milne, G.W., Pommier, Y., *J. Med. Chem.*, 40 (1997) 920.
18. Neamati, N., Mazumder, A., Zhao, H., Sunder, S., Burke, T.R., Jr., Schultz, R.J., Pommier, Y., *Antimicrob. Agents Chemother.*, 41 (1997) 385.
19. Neamati, N., Mazumder, A., Sunder, S., Owen, J.M., Schultz, R.J., Pommier, Y., *Antiviral Chem. Chemother.*, 8 (1997) 485.
20. Neamati, N., Turpin, J.A., Winslow, H.E., Christensen, J.L., Williamson, K., Orr, A., Rice, W.G., Pommier, Y., Garofalo, A., Brizzi, A., Campiani, G., Fiorini, I., Nacci, V.J., *Med. Chem.*, 42 (1999) 3334.
21. Cramer, R.D., III, Patterson, D.E., Bunce, J.D., *J. Am. Chem. Soc.*, 110 (1988) 5959.
22. Kulkarni, S.S., Kulkarni, V.M., *J. Med. Chem.*, 42 (1999) 373.
23. Gokhale, V.M., Kulkarni, V.M., *J. Med. Chem.*, 42 (1999) 5348.
24. Cho, S.J., Tropsha, A., *J. Med. Chem.*, 38 (1995) 1060.
25. Klebe, G., Abraham, U., Mietzner, T., *J. Med. Chem.*, 37 (1994) 4130.
26. Klebe, G., *Perspect Drug Discovery Des.*, 12 (1998) 87.
27. Bohm, M., Sturzebecher, J., Klebe, G., *J. Med. Chem.*, 42 (1999) 458.
28. Zhao, H., Neamati, N., Hong, H., Mazumder, A., Wang, S., Sunder, S., Milne, G.W.A., Pommier, Y., Burke, T.R., Jr., *J. Med. Chem.*, 40 (1997) 242.
29. Makhija, M.T., Kulkarni, V.M., *J. Chem. Inf. Comput. Sci.*, (2001) (appeared online in October 5<sup>th</sup> issue).
30. SYBYL *Molecular Modeling System*, Version 6.6; Tripos, Inc., St. Louis, MO 63144-2913.
31. Clark, M., Cramer, R.D., III, Van Opdenbosh, N.J., *Comput. Chem.*, 10 (1989) 982.
32. MOPAC 6.0 is available from Quantum Chemistry Program Exchange, Indiana University.
33. Stewart, J.J.P., *J. Comput.-Aided Mol. Des.*, 4 (1990) 1.
34. Kearsley, S.K., Smith, G.M., *Tetrahedron Comput. Methodol.*, 3 (1990) 615.
35. Wold, S., Albano, C., Dunn, W.J., III, Edlund, U., Esbensen, K., Geladi, P., Hellberg, S., Johansson, E., Lindberg, W., Sjostrom, M., *Multivariate Data Analysis in Chemistry. In CHEMOMETRICS: Mathematics and Statistics in Chemistry*; Kowalski, B., Ed.; Reidel: Dordrecht, The Netherlands, 1984.
36. Waller, C.L., Oprea, T.L., Giolitti, A., Marshall, G.R., *J. Med. Chem.*, 36 (1993) 4152.
37. Cramer, R.D., III, Bunce, J.D., Patterson, D.E., *Quant. Struct. Act. Relat.*, 7 (1988) 18.
38. Cerius2 3.5 is available from Molecular Simulations Inc., Scranton Road, San Diego, CA.
39. Raghavan, K., Buolamwini, J.K., Fesen, M.R., Pommier, Y., Kohn, K.W., Weinstein, J., *J. Med. Chem.*, 38 (1995) 890.
40. SYBYL 'Ligand-Based Design Manual' version 6.6; Tripos, Inc., St. Louis, MO.
41. Ouali, M., Laboulais, C., Leh, H., Gill, D., Desmaele, D., Mekouar, K., Zouhiri, F., Angelo, J., Auclair, C., Mouscadet, J.F., Bret, M.L., *J. Med. Chem.*, 43 (2000) 1949.
42. Ma, C.J., Dougherty, D.A., *Chem. Rev.*, 97 (1997) 1303.
43. Goldgur, Y., Craigie, R., Cohen, H.G., Fujiwara, T., Yoshinaga, T., Fujishita, T., Sugimoto, H., Endo, T., Murai, H., Davies, R.D., *Proc. Natl. Acad. Sci. USA*, 96 (1999) 13040.
44. Levit, M., Lifson, S., *J. Mol. Biol.*, 46 (1969) 269.
45. Gokhale, V.M., Kulkarni, V.M., *J. Comput.-Aided Mol. Des.*, 14 (2000) 495.
46. Orozco, M., Laughton, C.A., Herzyk, P., Neidle, S., *J. Biomol. Struct. Dyn.*, 8 (1990) 359.