



Molecular electrostatic potentials as input for the alignment of HIV-1 integrase inhibitors in 3D QSAR

Mahindra T. Makhija and Vithal M. Kulkarni*

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

Received 26 January 2001; accepted 26 July 2001

Key words: comparative molecular similarity indices analysis, HIV-1 integrase, molecular electrostatic potentials, partial least squares

Summary

Comparative molecular similarity indices analysis (CoMSIA), a three-dimensional quantitative structure activity relationship (3D QSAR) paradigm, was used to examine the correlations between the calculated physicochemical properties and the *in vitro* activities (3'-processing and 3'-strand transfer inhibition) of a series of human immunodeficiency virus type 1 (HIV-1) integrase inhibitors. The training set consisted of 34 molecules from five structurally diverse classes: salicylpyrazolinones, dioxepinones, coumarins, quinones, and benzoic hydrazides. The data set was aligned using extrema of molecular electrostatic potentials (MEPs). The predictive ability of the resultant model was evaluated using a test set comprised of 7 molecules belonging to a different structural class of thiazepinediones. A CoMSIA model using an MEP-based alignment showed considerable internal as well external predictive ability ($r_{cv}^2 = 0.821$, $r_{pred.}^2 = 0.608$ for 3'-processing; and $r_{cv}^2 = 0.759$, $r_{pred.}^2 = 0.660$ for 3'-strand transfer).

Introduction

The pandemic spread of acquired immunodeficiency syndrome (AIDS) has promoted an unprecedented scientific and clinical effort to understand and combat this lethal disease. Human immunodeficiency virus (HIV) is an etiologic agent of AIDS, which expresses its effects through the genetic direction of viral polypeptides prepared by the host. Several key enzymes in the replication cycle of the HIV can be targeted for chemotherapeutic intervention, most notably, reverse transcriptase and HIV protease [1]. Combination antiviral therapy with HIV protease and reverse transcriptase inhibitors has demonstrated the potential therapeutic efficacy of antiviral therapy for treatment of AIDS [2]. However, the long-lived nature of the infection and the genetic plasticity of the virus have made it apparent that new antiretroviral agents

are required to deal with the appearance and spread of resistance.

HIV integrase is an attractive target for antivirals because it is essential for HIV replication and, unlike protease and reverse transcriptase, there are no known counterparts in the host cell [3]. Furthermore, integrase uses a single active site to accommodate two different configurations of DNA substrates, which may constrain the ability of HIV to develop drug resistance to integrase inhibitors [4–6].

Retroviruses encode the integrase protein at the 3'-end of the *pol* gene. This enzyme, an HIV protease cleavage product of the *gag-pol* fusion protein precursor, catalyzes the integration of a double stranded DNA copy of the RNA genome, synthesized by the reverse transcriptase, into a host chromosome in a two step reaction. First, integrase cleaves the last two nucleotides from each 3'-end of the linear viral DNA, leaving the terminal dinucleotide CA-3'-OH. This activity is referred to as 3'-processing or dinucleotide cleavage. Secondly, after transport to the nucleus as

*To whom correspondence should be addressed. E-mail: vithal@biogate.com

a nucleoprotein complex, integrase catalyzes a DNA strand transfer reaction involving a nucleophilic attack from the cleaved 3'-ends to a host chromosome. This process is referred to as strand transfer. Finally, the viral 5'-ends are processed, and the gaps between the viral 5'- and target 3'-ends are repaired [7, 8].

Although several classes of integrase inhibitors have been reported to date, none of them has proven to be highly selective and useable for therapeutic development [9, 10]. For example, a majority of catechol-containing compounds have been demonstrated to possess potency against the 3'-processing and 3'-strand transfer reactions catalyzed by IN [10]. However, despite such promising initial results, compounds containing this moiety can also cross-link protein [11], chelate metal [12, 13], and thus generally lack the desired selectivity [9, 10]. In contrast, inhibitors, which lack the catechol substructure, are excellent leads to develop a selective potent integrase inhibitor. Pommier et al. have identified several such non catechol containing compounds like salicylpyrazolinones [12], dioxepinones [13], coumarins [14], quinones [15], benzoic hydrazides [16], thiazepinediones [17], etc., that inhibit integrase function at low micromolar concentrations.

To further explore the relationship between the structures of the aforementioned classes of compounds and their integrase inhibitory activities, three-dimensional quantitative structure activity relationship (3D QSAR) studies using comparative molecular similarity indices analysis (CoMSIA) have been performed. The CoMSIA method of 3D QSAR was introduced by Klebe in 1994, in which molecular similarity is evaluated in space [18]. Similarity is expressed in terms of different physicochemical properties: steric occupancy, partial atomic charges, local hydrophobicity, and hydrogen-bond donor and acceptor properties. Using a common probe atom, similarity indices are calculated for a data set of prealigned molecules at regularly spaced grid points. In determining such similarities, the mutual distances between the probe and the atoms of the molecules of the data set are considered. For this distance dependence, a Gaussian type functional form is selected that avoids singularities at the atomic positions and requires no arbitrary definition of cutoff limits. The similarity indices calculated at all grid points (inside or outside the molecular van der Waal's surface) are useful descriptors for subsequent analysis by partial least squares (PLS). By default, CoMSIA uses five different aforementioned property fields to partition them into spatial loca-

tions where they play a decisive role in determining biological activity [19–21].

The CoMSIA method of 3D QSAR requires molecules to be prealigned for calculation of fields. Molecular alignment can be relatively straightforward if a set of rigid structural analogues is to be analyzed but becomes increasingly problematic as the diversity of the dataset increases, even if no account is taken of conformational flexibility [22]. One such class is that of integrase inhibitors which present a tremendous diversity of molecular structure. Recently, an alignment-free method of 3D QSAR eigen value analysis (EVA) was performed on integrase inhibitors [23]. But, alignment based methods like CoMFA [24–26] and CoMSIA have an advantage over EVA that 3D visualization of fields in the form of isocontour plots is possible which helps in identifying the regions of space that have positive or negative correlations with the biological activity. To align the data set of integrase inhibitors, molecular electrostatic potentials (MEP) were used.

Molecular electrostatic potentials (MEPs) [27–29]

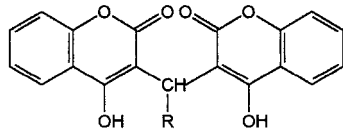
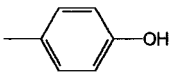
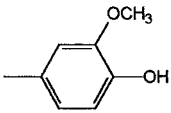
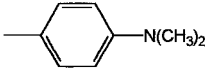
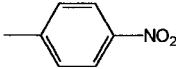
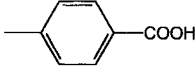
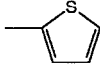
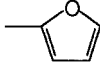
The initial step in many biological processes such as drug-receptor or enzyme-substrate interactions is molecular recognition, in which the receptor recognizes key features of the approaching drug. This recognition, which precedes binding, is believed to occur when two molecules involved in the process are at a relatively large separation. The electrostatic potential has proved to be very useful in rationalizing the interactions between molecules at large separations and for molecular recognition processes. Loosely speaking two molecules first see each other by means of the electrostatic potential. The MEP is a physically observable quantity (e.g., by X-ray diffraction) and can be derived directly from the wave function. MEP is expressed as equipotential contours and positions of minimum potentials. The nuclear and electronic charge distribution of the molecule creates an electrostatic potential, which will interact with a point charge or the electron density of another molecule.

The electrostatic potential at any point r in space can be expressed as:

$$V(r) = \sum_A (Z_A / (R_A - r)) - \int (\rho(r') \, dr' / (r' - r)),$$

where Z is the charge at nucleus A located at R , $\rho(r)$ is the electronic density function. The first term represents the nuclear contribution and the second term is

Table 1. Structures and activities of coumarins used in the training set (from ref. 14)

			
Sr.No.	R	3' -Processing IC ₅₀ (mM)	3' -Strand transfer IC ₅₀ (mM)
1		0.134	0.074
2		0.187	0.092
3		0.094	0.069
4		0.054	0.018
5		0.057	0.051
6		0.105	0.148
7		0.054	0.122

the electronic contribution. The electron density used to calculate the potential can be obtained from either an *ab initio* or semiempirical wave function and is approximate, as is the potential. Those regions that have high nuclear contribution will yield positive MEP, corresponding to repulsive interaction energies with point positive charges, and those with higher electron contributions will yield negative MEP, corresponding to attractive energies. The drug-receptor-type recognition process should be one in which regions with opposite MEPs match. In the present study, electrostatic potentials of all molecules were calculated and displayed and the most electropositive and most electronegative regions were identified for all molecules. Using these extrema, i.e., most electropositive and most electronegative regions, all molecules were superimposed on the MEP for the most potent example considered (**10**; see Table 2).

Materials and methods

Data set for analysis

Published *in vitro* biological data for both 3'-processing and 3'-endjoining (strand transfer) on a series of HIV integrase inhibitors were used for this study [12–17]. The structures and biological activities of the 34 molecules constituting the training set in the QSAR analyses are given in Table 1–5, respectively.

The test set consists of 7 molecules (Table 6). These compounds belong to an entirely different structural class and were selected to remove any bias in testing the predictivity of the models.

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, 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]. MEPs were displayed using MOLCAD in SYBYL 6.6.

Molecular conformation

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 energy gradient convergence criterion. Partial atomic charges were computed by a semiempirical molecular orbital method using the MOPAC 6.0 program [32]. The charges were computed using the PM3 model Hamiltonian (keywords : 1SCF, RHF, MMOK) [33].

Using the systematic search protocol, rotatable bonds in compounds were searched from 0 to 360° in 10° increments. The low-energy conformations of these compounds thus obtained, were minimized using Tripos forcefield and subsequently used in the analysis.

CoMSIA Analyses

CoMSIA calculates similarity indices at the intersections of a surrounding lattice. The similarity index A_F [34] for a molecule j with i atoms at the grid point q is determined as follows:

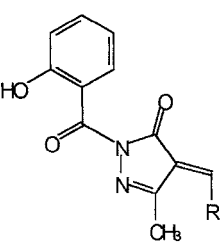
$$A_{F,k}^q(j) = -\sum_i \omega_{\text{probe},k} \omega_{ik} e_{iq}^{-\alpha r^2}$$

where ω_{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. Five physicochemical properties k (steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor) were evaluated, using a common probe atom with 1 Å radius and charge, and hydrophobicity, and hydrogen-bond property of +1. 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 ' α ' 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. 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 [35].

Table 2. Structures and activities of salicylpyrazolinones used in the training set (from ref. 12)



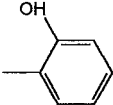
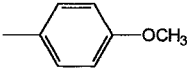
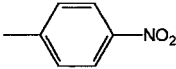
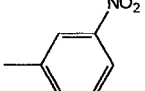
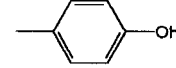
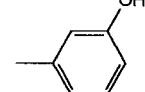
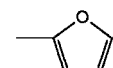
Sr.No.	R	3'-Processing IC ₅₀ (mM)	3'-Strand transfer IC ₅₀ (mM)
8		0.0006	0.0028
9		0.0009	0.0006
10		0.0008	0.0006
11		0.0014	0.0026
12		0.0006	0.0009
13		0.0009	0.0074
14		0.0027	0.0020

Table 3. Structures and activities of quinones used in the training set (from ref. 15)

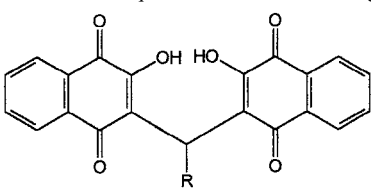
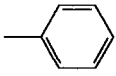
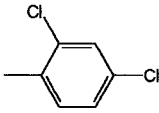
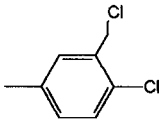
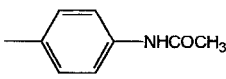

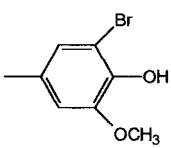
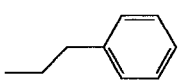
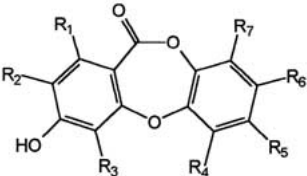
			
Sr.No.	R	3' -Processing IC ₅₀ (mM)	3' -Strand transfer IC ₅₀ (mM)
15		0.068	0.048
16		0.037	0.040
17		0.090	0.052
18		0.086	0.078
19		0.092	0.060
20		0.032	0.020
21		0.083	0.090

Table 4. Structures and activities of dioxepinones used in the training set (from ref. 13)



No.	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	R ₇	IC ₅₀ ^a	IC ₅₀ ^b
22	Me	H	CHO	Me	CO ₂ H	OH	Me	0.0046	0.0065
23	Me	H	CHO	Me	CO ₂ Me	OH	Me	0.0054	0.0044
24	Me	H	CHO	Me	CO ₂ H	OH	CH ₂ OCOCH:CHCO ₂ H	0.0049	0.0046
25	CH ₂ COC ₅ H ₁₁	H	H	n-C ₅ H ₁₁	CO ₂ H	OH	H	0.0385	0.0309
26	COC ₄ H ₉	H	H	n-C ₅ H ₁₁	H	OH	H	0.0510	0.0336
27	Me	H	CHO	CO ₂ H	H	OMe	Me	0.0160	0.0088
28	Me	Cl	CHO	Me	H	OMe	Me	0.0024	0.0021

^aIC₅₀ (mM) values for 3'-processing activity.^bIC₅₀ (mM) values for 3'-strand transfer activity.

PLS analyses were performed following the CoMFA standard implementation in SYBYL. To check statistical significance of the models, cross-validations were done by means of the 'leave-one-out' (LOO) procedure using the enhanced version of PLS, the SAMPLS method [36]. 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 individual compound based on the data available from other compounds in the set

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

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. Additionally to perform an even more rigorous statistical test, sev-

eral 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.

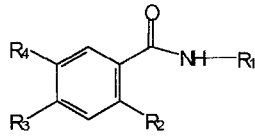
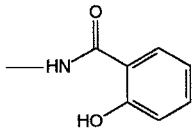
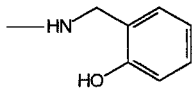

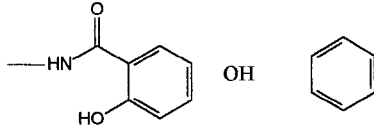

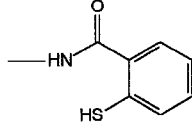

A common test to check the consistency of the models is to scramble the biological data and repeat the model derivation process, allowing detection of possible chance correlations. After randomizing our data set in several distinct ways, in all cases we only observed very low or negative r_{cv}^2 values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

$$r_{\text{pred}}^2 = (\text{SSD} - \text{PRESS}) / \text{SSD},$$

where SSD 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. 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, 38].

Table 5. Structures and activities of benzoic hydrazides used in the training set (from ref. 16)

						
Sr.No.	R ₁	R ₂	R ₃	R ₄	3'-Processing IC ₅₀ (mM)	3'-Strand transfer IC ₅₀ (mM)
29		OH	H	H	0.0020	0.0007
30		OH	H	H	0.0067	0.0052
31		OH	H	H	0.0800	0.0380
32 ^a		OH		H	0.0023	0.0011
33		OH	H	H	0.0091	0.0058
34		SH	H	H	0.0748	0.0537

^a Fused ring system (naphthalene derivative).

Results and discussion

The results of CoMSIA analyses for both 3'-processing and 3'-strand transfer are summarized in Table 7. CoMSIA analysis when performed on data set per se, without any alignment yielded a correlation with poor r_{cv}^2 of 0.007 and -0.061 using 3 and 2 principal components for 3'-processing and strand transfer respectively. The conventional r^2 for these analyses were 0.954 and 0.815, while predictive r^2 were 0.229 and 0.309 for 3'-processing and strand transfer respectively. Thus, in the absence of any alignment method,

models were found to have poor internal as well external predictive abilities. High s_{press} values of 0.852 and 0.798 for 3'-processing and strand transfer confirm the poor predictive abilities of these models.

When the data set was aligned using extrema of MEPs, leave one out r_{cv}^2 improved to a considerable extent (0.821 for 3'-processing 0.759 for 3'-strand transfer). Both models required five principal components to explain the variances in biological activity. The conventional r^2 for these analyses were 0.980 and 0.963 for 3'-processing and strand transfer. These

Table 6. Structures and activities of thiazepinediones used in the test set (from ref. 17)

Sr.No.	R ₁	R ₂	-X-Y-	3'-Processing IC ₅₀ (mM)	3'-Strand transfer IC ₅₀ (mM)
35	H	Cl	-S-CH ₂ -	0.128	0.090
36	H	Br	-S-CH ₂ -	0.058	0.048
37	H	CH ₃	-S-CH ₂ -	0.064	0.055
38	H	OCH ₃	-CH ₂ -S-	0.215	0.200
39	OCH ₃	OCH ₃	-CH ₂ -S-	0.650	0.331
40 ^a			-S-CH ₂ -	0.040	0.047
41 ^a			-CH ₂ -S-	0.092	0.100

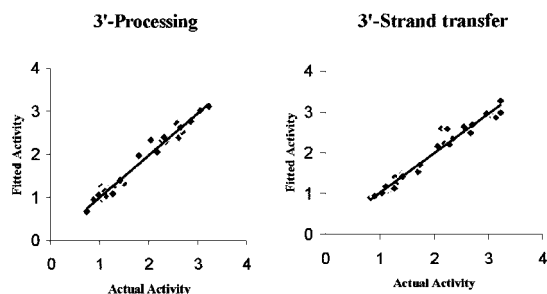
^a Fused ring system (naphthalene derivative).

Figure 1. Fitted versus actual activity values for CoMSIA analysis of the training set.

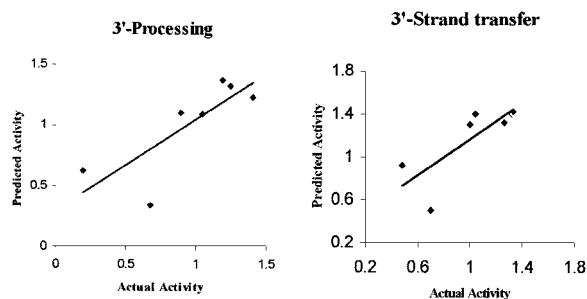


Figure 2. Predicted versus actual activity values for CoMSIA analysis of the test set.

models showed good external predictivities ($r_{\text{pred}}^2 = 0.608$ for 3'-processing and 0.660 for 3'-strand transfer) for the test set which belongs to an entirely different structural class. Thus, the models are truly predictive and unbiased. The high bootstrapped r^2 values and low standard deviations ($r_{\text{bs}}^2 = 0.989$ and $\text{SD} = 0.005$ for 3'-processing; $r_{\text{bs}}^2 = 0.979$ and $\text{SD} =$

0.010 for 3'-strand transfer) suggest a high degree of confidence in these models.

The stability and the consistency of the final models were further checked by randomization of biological activities and with cross-validations using five groups. The results of these analyses are summarized in Table 8. Negative r_{cv}^2 values obtained by scrambling

Table 7. Summary of results from the CoMSIA analyses

	Unaligned		MEP-aligned	
	3'-Processing	3'-Strand transfer	3'-Processing	3'-Strand transfer
r_{cv}^2	0.007	-0.06	0.821	0.759
s_{press}	0.852	0.798	0.375	0.400
$r_{conv.}^2$	0.954	0.815	0.980	0.963
SE	0.183	0.333	0.124	0.157
Components	3	2	5	5
F-value	208.03	68.153	279.92	144.43
$Pr^2 = 0$	0.000	0.000	0.000	0.000
Contribution:				
Steric	0.134	0.115	0.155	0.141
Electrostatic	0.206	0.198	0.200	0.182
Hydrophobic	0.263	0.307	0.239	0.263
H-bond donor	0.216	0.193	0.238	0.262
H-bond acceptor	0.182	0.187	0.167	0.152
$r_{pred.}^2$	0.229	0.309	0.608	0.660
r_{bs}^{2a}	0.962	0.875	0.989	0.979
SD ^a	0.014	0.040	0.005	0.010

^aResults from 1000 runs of bootstrapped analysis.

the activity data are consistent with the stability of the models.

In CoMSIA five different fields are used for describing the physicochemical properties of the molecules. Since functional group replacement among the considered ligands can affect more than one field with similar consequences (e.g., removal of a group affects both steric and electrostatic properties as they are partially correlated or replacement of charged groups influences coherently electrostatic and hydrophobic properties), individual fields were correlated with two activities to assess their individual contributions to the derived models. The results of these analyses are summarized in Table 9 and 10 for 3'-processing and 3'-strand transfer respectively. All five fields were found to contribute significantly.

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 Figures 1 and 2, respectively. The unaligned data set is shown in Figures 3 and 4 shows an overlay of exemplars from each class after aligning with MEPs. The observed and calculated pIC₅₀ values are given in Table 11 for the training set and observed and predicted pIC₅₀ values for the test set are given in Table 12.

All other five classes of compounds used in the training set were individually selected as test set and remaining compounds for the training set. The model

Table 8. Results of analyses with randomized biological activities and cross-validation using five groups

	3'-Processing		3'-Strand transfer	
	r_{cv}^{2a}	r_{cv}^{2b}	r_{cv}^{2a}	r_{cv}^{2b}
Mean	-0.21	0.79	-0.32	0.73
Std.Dev.	0.13	0.04	0.11	0.04
High	0.03	0.85	-0.07	0.81
Low	-0.52	0.70	-0.51	0.61

^aCross-validated r^2 with randomized biological activity average of 25 runs.

^bCross-validation using five groups with optimum number of components average of 25 runs.

derivation was repeated and internal as well as external predictivities of these models were checked. The results are impressive and are summarized in Table 13 and 14 for 3'-processing and 3'-strand transfer respectively. Thus, the model is predictive and unbiased.

Graphical interpretation of results

In order to visualize the derived 3D QSAR model, CoMSIA contour maps were generated by interpolating the products between the 3D QSAR coefficients and their associated standard deviations:

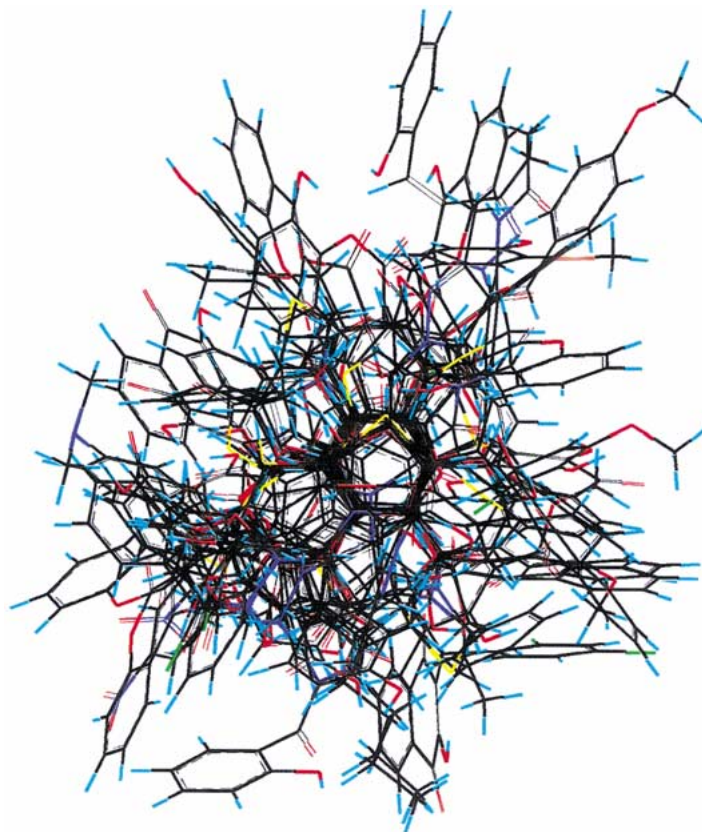


Figure 3. Unaligned molecules in the data set.

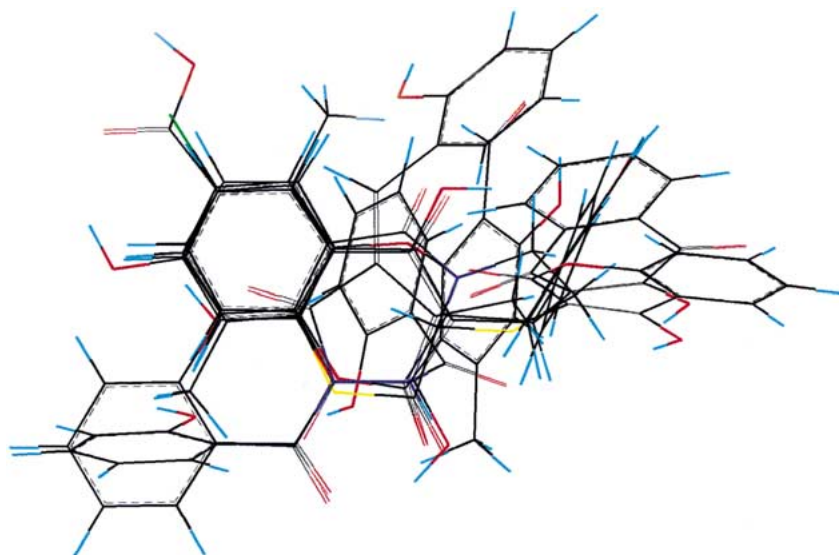


Figure 4. Molecules (1, 8, 15, 22, 29, and 35) from each class aligned using MEPs.

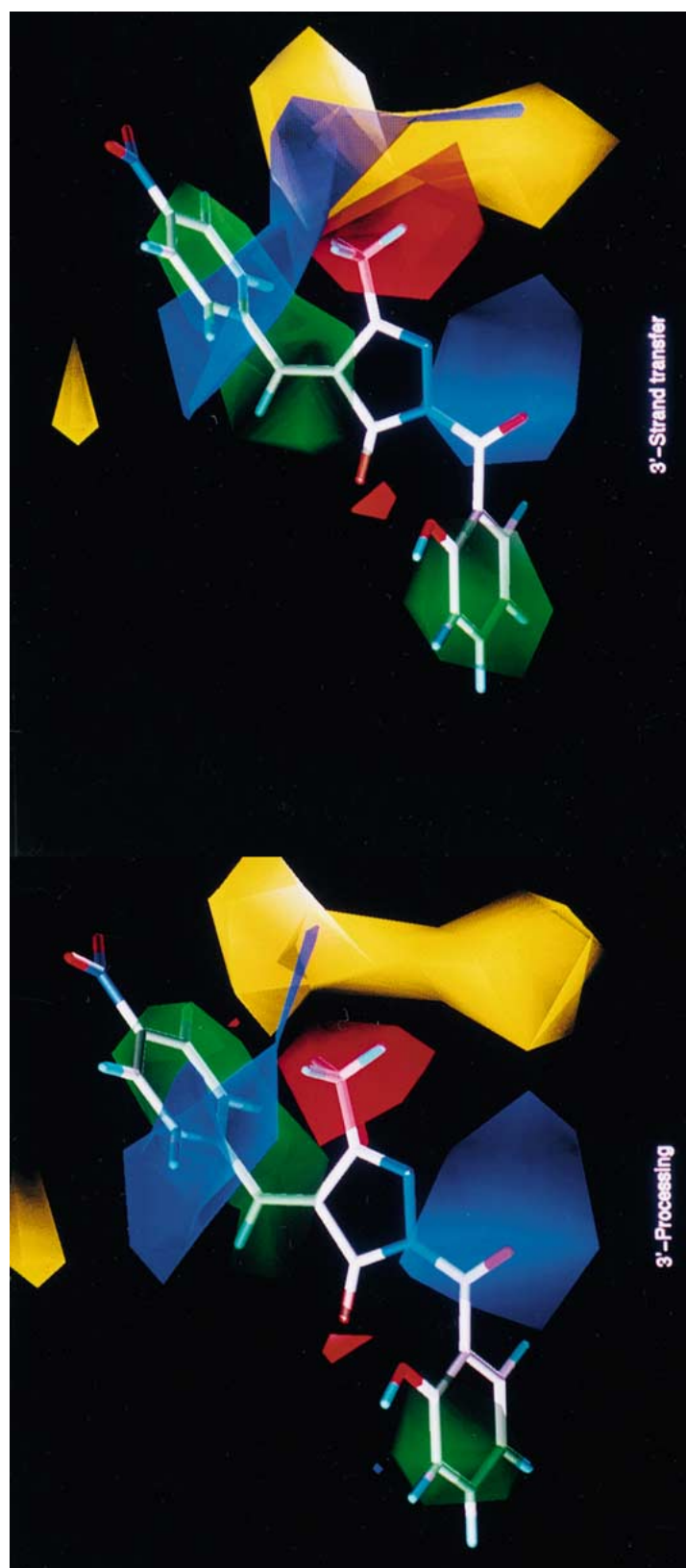


Figure 5. Steric and electrostatic maps from the CoMSIA model with compound 10 as a reference structure. (a) 3'-Processing. (b) 3'-Strand transfer. Green contours (80% contribution) enclose areas where steric bulk will enhance affinity and yellow contours (20% contribution) highlight areas which should be kept unoccupied; otherwise binding affinity will decrease. 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.

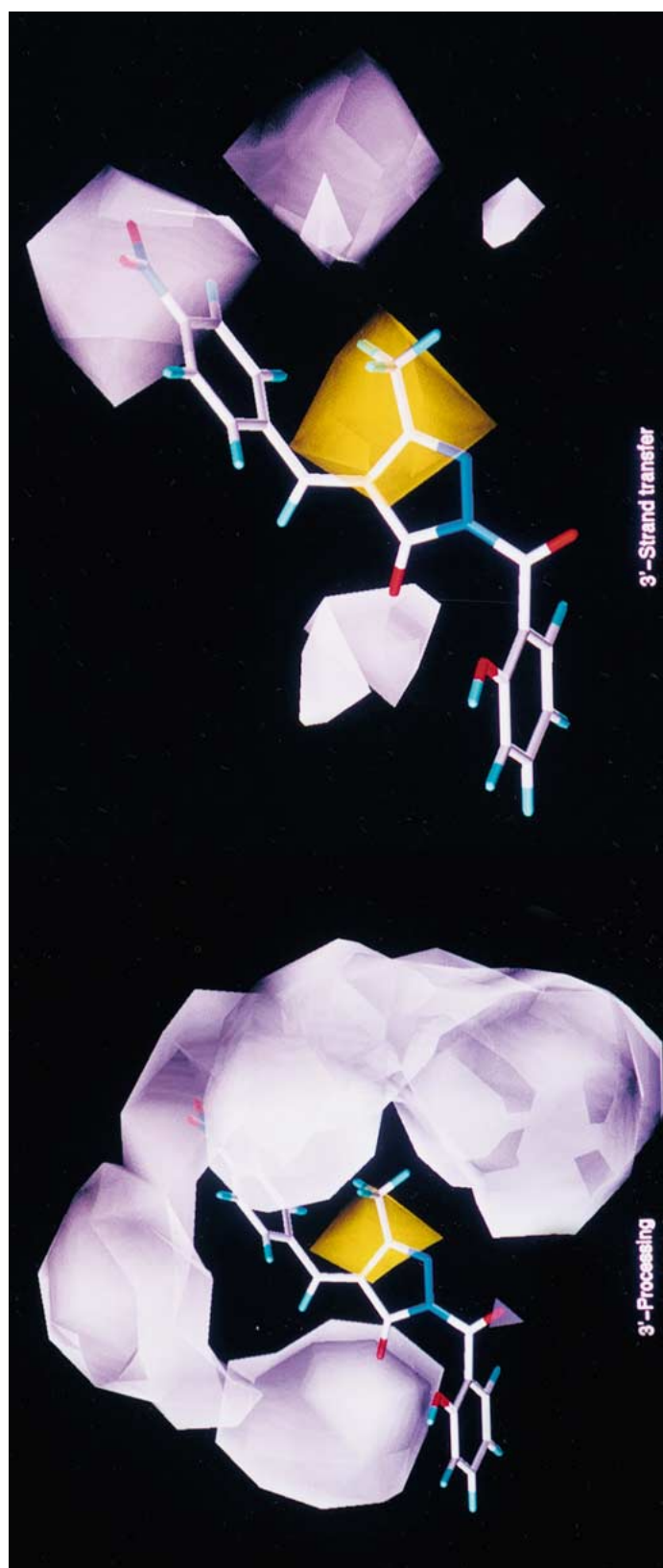


Figure 6. Hydrophobic maps from the CoMSIA model with compound 10 as a reference structure. (a) 3'-Processing. Yellow contours (80% contribution) enclose areas where hydrophobicity will enhance affinity and white contours (20% contribution) highlight areas where increase in hydrophobicity will decrease the binding affinity.

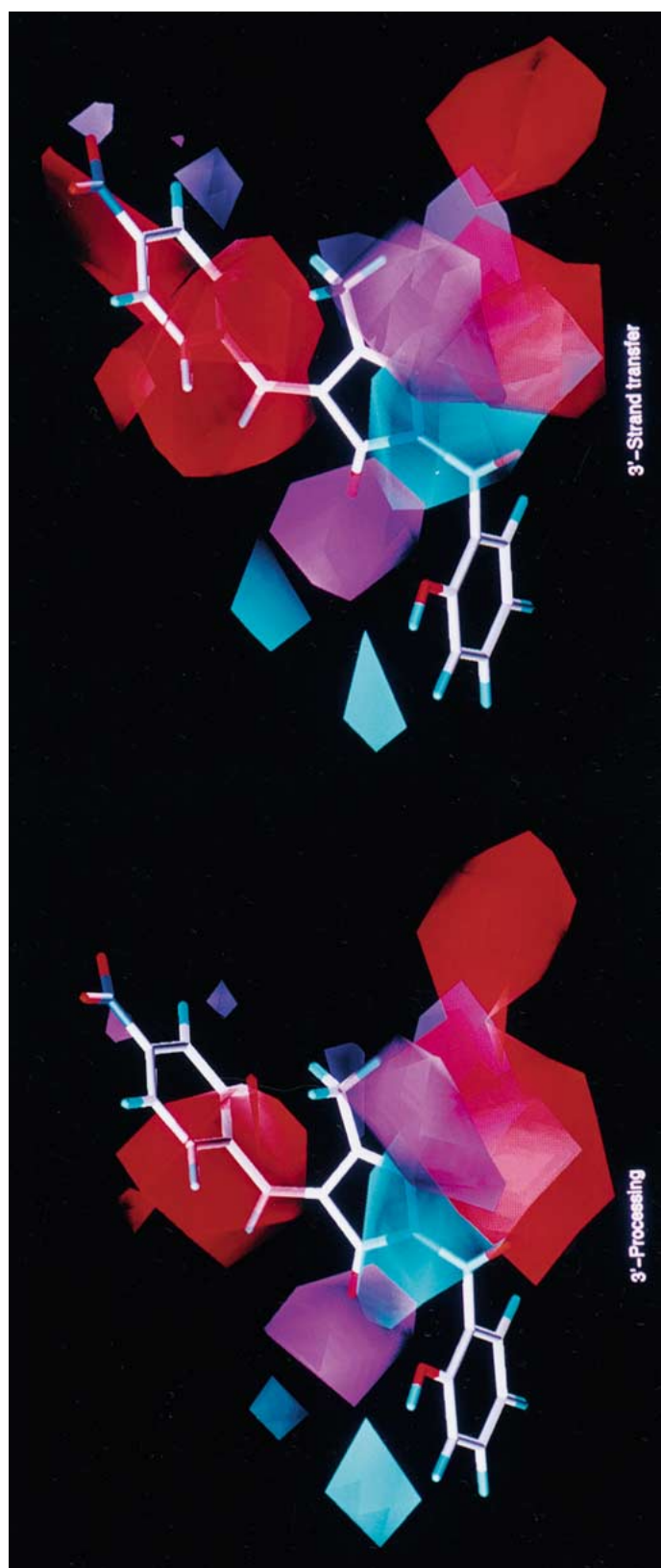


Figure 7. Hydrogen-bond Donor and Acceptor fields from the CoMSIA model with compound 10 as a reference structure. (a) 3'-Processing. (b) 3'-Strand transfer. Cyan contours (80% contribution) enclose areas where donor fields will enhance affinity and purple contours (20% contribution) highlight areas where donor fields are not important. Magenta isopleths (80% contribution) encompass regions where acceptor fields are important and red contours (20% contribution) represent regions where acceptor fields do not play a role in HIV-1 integrase inhibition.

Table 9. Summary of MEP-aligned CoMSIA results for 3'-processing using individual fields

	Steric	Electrostatic	Hydrophobic	H-bond donor	H-bond acceptor
r_{cv}^2	0.886	0.702	0.797	0.699	0.850
s_{press}	0.299	0.475	0.399	0.469	0.349
$r_{conv.}^2$	0.952	0.931	0.966	0.903	0.972
SE	0.194	0.228	0.163	0.266	0.150
comp.	5	4	5	3	6
F-value	111.0	98.36	160.1	93.55	158.6

Table 10. Summary of MEP-aligned CoMSIA results for 3'-strand transfer using individual fields

	Steric	Electrostatic	Hydrophobic	H-bond donor	H-bond acceptor
r_{cv}^2	0.773	0.531	0.701	0.650	0.722
s_{press}	0.388	0.530	0.446	0.474	0.430
$r_{conv.}^2$	0.899	0.750	0.925	0.908	0.921
SE	0.259	0.387	0.223	0.243	0.229
comp.	5	2	5	4	5
F-value	49.89	46.49	69.13	71.64	65.08

(a) *Steric and electrostatic fields.* Figure 5 shows a stereoview of the CoMSIA steric and electrostatic contour map from the analysis based on MEP alignment using compound 10 as a reference structure. 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); and 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). There is not much difference between the steric and electrostatic contours for 3'-processing and 3'-strand transfer activities, indicating similar steric and electrostatic requirements in terms of substituents on inhibitors for inhibition of both these activities. The presence of green contour in the vicinity of two phenyl rings indicates that steric bulk would be favorable at these positions for integrase inhibition. Blue polyhedra near the hydroxyl group and the carbonyl functions of the pyrazole ring indicates the involvement of these groups in electrostatic interactions with integrase.

(b) *Hydrophobic fields.* Figure 6 shows a stereoview of the hydrophobic contour map using compound 10 as a reference structure. In this figure, white contours (20% contribution) represent regions where hydrophobicity of the substituent or placement of hy-

drophobic groups diminish the integrase inhibitory activity, while yellow contours (80% contribution) represent areas where substitution of hydrophobic groups would be favorable for activity. From Figure 6, it seems that the requirements for hydrophobic fields is different for both the activities. In case of 3'-processing, the hydrophobically disfavored region surrounds almost entire molecule, leaving only small yellow patch where methyl group is accommodated; whereas for 3'-strand transfer activity, the disfavored region for hydrophobic interactions surrounds the terminal phenyl ring and some parts of the pyrazolone nucleus. This difference in terms of requirements for hydrophobic fields may help in understanding the mechanism of 3'-processing and strand transfer at the molecular level. Thus, binding of these inhibitors to integrase seems mainly enthalpic in nature and during binding the areas that get desolvated are not hydrophobic but are rather polar.

(c) *Hydrogen-bond donor and acceptor fields.* Figure 7 shows a stereoview of hydrogen-bond (H-bond) donor and acceptor fields using compound 10 as a reference structure. In this figure, cyan and magenta contours represent regions where H-bond donor and acceptor fields, respectively favor activity (80% contribution); while purple and red areas indicate regions where H-bond donor and acceptor fields respectively

Table 11. Observed versus calculated pIC₅₀ values using MEP alignment

Compound	3'-processing		3'-strand transfer	
	Observed pIC ₅₀	Calculated pIC ₅₀	Observed pIC ₅₀	Calculated pIC ₅₀
1	0.871	0.961	1.127	1.116
2	0.726	0.685	1.036	1.067
3	1.026	1.029	1.161	1.084
4	1.267	1.246	1.744	1.711
5	1.244	1.135	1.288	1.255
6	0.978	1.059	0.828	0.911
7	1.263	1.088	0.911	0.951
8	3.221	3.127	2.552	2.644
9	3.045	3.031	3.221	2.980
10	3.096	3.055	3.221	3.272
11	2.853	2.783	2.585	2.518
12	3.221	3.179	3.045	2.910
13	3.045	3.089	2.130	2.604
14	2.568	2.729	2.698	2.694
15	1.167	1.227	1.318	1.436
16	1.431	1.393	1.397	1.500
17	1.045	1.258	1.283	1.406
18	1.065	1.178	1.107	1.181
19	1.036	0.975	1.221	1.118
20	1.494	1.338	1.698	1.539
21	1.080	1.048	1.045	1.022
22	2.337	2.309	2.187	2.231
23	2.267	2.334	2.356	2.305
24	2.309	2.412	2.337	2.346
25	1.414	1.412	1.510	1.412
26	1.292	1.282	1.473	1.424
27	1.794	1.990	2.055	2.172
28	2.610	2.404	2.667	2.490
29	2.684	2.533	3.136	2.873
30	2.173	2.059	2.283	2.210
31	1.096	1.152	1.420	1.429
32	2.638	2.641	2.958	2.966
33	2.040	2.341	2.236	2.589
34	1.126	1.042	1.269	1.135

Table 12. Observed versus predicted pIC₅₀ values for test set molecules.

Compound	3'-processing		3'-strand transfer	
	Observed pIC ₅₀	Predicted pIC ₅₀	Observed pIC ₅₀	Predicted pIC ₅₀
35	0.892	1.100	1.045	1.400
36	1.236	1.320	1.318	1.400
37	1.193	1.370	1.259	1.320
38	0.667	0.340	0.698	0.500
39	0.187	0.630	0.479	0.923
40	1.397	1.230	1.327	1.420
41	1.036	1.090	1.000	1.300

Table 13. Summary of CoMSIA results using MEP based alignment with different series as test sets for 3'-processing

Sr. No.	Test set	r_{cv}^2	Comp.	r_{conv}^2	F-value	r_{pred}^2
1	Coumarins (Table 1)	0.815	6	0.984	276.89	0.603
2	Salicylpyrazolinones (Table 2)	0.644	6	0.968	136.07	0.569
3	Quinones (Table 3)	0.822	5	0.983	316.14	0.921
4	Dioxepinones (Table 4)	0.841	4	0.956	159.05	0.530
5	Benzoic hydrazides (Table 5)	0.868	6	0.985	301.05	0.601

Table 14. Summary of CoMSIA results using MEP based alignment with different series as test sets for 3'-strand transfer

Sr. No.	Test set	r_{cv}^2	Comp.	r_{conv}^2	F-value	r_{pred}^2
1	Coumarins (Table 1)	0.772	6	0.976	184.91	0.588
2	Salicylpyrazolinones (Table 2)	0.727	6	0.973	160.28	0.636
3	Quinones (Table 3)	0.781	6	0.981	227.98	0.931
4	Dioxepinones (Table 4)	0.814	6	0.977	191.33	0.576
5	Benzoic hydrazides (Table 5)	0.854	6	0.984	282.68	0.661

are disfavored for activity. In the vicinity of the salicylic hydroxyl group, there is a cyan isopleth, which shows its H-bond donor character. Presence of magenta contour near the hydrazide function and NO₂ group indicates that these groups may accept a proton from an enzyme to form a H-bond. Because H-bonding capabilities are also reflected in electrostatic fields, similar features appear in the electrostatic contour maps as well.

Recently, we also performed genetic algorithm based QSAR studies on this data set with 32 descriptors calculated using Cerius2. The results indicate that spatial, thermodynamic and shape descriptors contribute significantly to 3'-processing inhibition. For 3'-strand transfer inhibition, apart from spatial and thermodynamic descriptors, structural parameters such as number of H bond Donor groups is also important [39].

Conclusions

The CoMSIA method has been applied successfully to rationalize the HIV integrase inhibitory activity of a diverse set of 41 inhibitors in terms of five different molecular fields. This model was developed from an alignment strategy based on MEPs, which takes into consideration the process of molecular recogni-

tion that precedes binding. The 3D QSAR culminating from the training set yielded a regression equation with a high degree of statistical significance and performed exceptionally well in predicting the biological activity of the test set, which belongs to an entirely different structural class of thiazolothiazepines. The contour diagrams obtained for the CoMSIA field contributions can be mapped back onto structural features accounting for the activity trends among the inhibitors. On the basis of the spatial arrangement of the field contributions, novel molecules can be designed that are predicted to possess improved integrase inhibitory activity.

Acknowledgements

The authors gratefully acknowledge support 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. and Kulkarni, V.M., *Pharmacy Pharmacology Commun.*, 4 (1998) 365.
2. De Clercq, E., *J. Med. Chem.*, 38 (1995) 2491.

3. Goldgur, Y., Craigie, R., Cohen, H. G., Fujiwara, T., Yoshinaga, T., Fujishita, T., Sugimoto, H., Endo, T., Murai, H. and Davies, R. D., *Proc. Natl. Acad. Sci. USA*, 96 (1999) 13040.
4. Sakai, H., Kawamura, M., Sakuragi, J., Sakuragi, S., Shibata, R., Ishimoto, A., Ono, H., Ueda, S. and Adachi, A., *J. Virol.*, 7 (1993) 1169.
5. Taddeo, B., Haseltine, W.A. and Farnet, C.M., *J. Virol.*, 68 (1994) 8401.
6. Engelman, A., Englund, G., Orenstein, J.M., Martin, M.A. and Craigie, R., *J. Virol.*, 69 (1995) 2729.
7. Goff, S.P., *Annu. Rev. Genet.*, 26 (1992) 527.
8. Craigie, R., *Trends Genet.*, 8 (1992) 187.
9. Pommier, Y., Pilon, A., Bajaj, K., Mazumder, A. and Neamati, N., *Antiviral Chem. Chemother.*, 8 (1997) 483.
10. Neamati, N., Sunder, S. and Pommier, Y., *Drug Discovery Today*, 2 (1997) 487.
11. Stanwell, C., Ye, B., Yuspa, S.H. and Burke, T.R., Jr., *Biochem. Pharmacol.*, 52 (1996) 475.
12. Neamati, N., Hong, H., Owen, J. M., Sunder, S., Winslow, H. E., Christensen, J. L., Zhao, H., Burke, T. R., Jr, Milne, G. W. A., Pommier, Y., *J. Med. Chem.*, 41 (1998) 3202.
13. Neamati, N., Hong, H., Mazumder, A., Wang, S., Sunder, S., Nicklaus, M.C., Milne, G.W.A., Proksa, B. and Pommier, Y., *J. Med. Chem.*, 40 (1997) 942.
14. Zhao, H., Neamati, N., Hong, H., Mazumder, A., Wang, S., Sunder, S., Milne, G.W. A., Pommier, Y. and Burke, T.R., Jr., *J. Med. Chem.*, 40 (1997) 242.
15. Mazumder, A., Wang, S., Neamati, N., Nicklaus, M., Sunder, S., Chen, J., Milne, G.W.A., Rice, W.G., Burke, T.R., Jr and Pommier, Y., *J. Med. Chem.*, 39 (1996) 2472.
16. Zhao, H., Neamati, N., Sunder, S., Hong, H., Wang, S., Milne, G.W.A., Pommier, Y. and Burke, T.R., Jr., *J. Med. Chem.*, 40 (1997) 937.
17. 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.I. Fiorini and Nacci, V., *J. Med. Chem.*, 42 (1999) 3334.
18. Klebe, G., Abraham, U. and Mietzner, T., *J. Med. Chem.*, 37 (1994) 4130.
19. Klebe, G., *Perspect. Drug Discovery Des.*, 12 (1998) 87.
20. Bohm, M., Sturzebecher, J. and Klebe, G., *J. Med. Chem.*, 42 (1999) 458.
21. Makhija, M.T. and Kulkarni, V.M., *J. Comput. Aid. Mol. Des.* (Manuscript submitted).
22. Talele, T.T., Kulkarni, S.S. and Kulkarni, V.M., *J. Chem. Inf. Comput. Sci.*, 39 (1999) 958.
23. Makhija, M.T. and Kulkarni, V.M., *J. Chem. Inf. Comput. Sci.* (in press).
24. Cramer, R.D., III, Patterson, D.E. and Bunce, J.D., *J. Am. Chem. Soc.*, 110 (1988) 5959.
25. Kulkarni, S.S. and Kulkarni, V.M., *J. Med. Chem.*, 42 (1999) 373.
26. Gokhale, V.M. and Kulkarni, V.M., *J. Med. Chem.*, 42 (1999) 5348.
27. Sjoberg, P. and Politzer, P., *J. Phys. Chem.*, 94 (1990) 3959.
28. Kosov, D.S. and Popelier, P.L.A., *J. Phys. Chem. A*, 104 (2000) 7339.
29. Kocjan, D., Hodoscek, M. and Hadzi, D., *J. Med. Chem.*, 29 (1986) 1418.
30. SYBYL Molecular Modeling System, version 6.6; Tripos, Inc., St. Louis, MO 63144-2913.
31. Clark, M., Cramer, R.D., III and 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. Aid. Mol. Des.*, 4 (1990) 1.
34. Kearsley, S.K. and 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. and Sjostrom, M., in Kowalski, B. (Ed.) *CHEMOMETRICS: Mathematics and Statistics in Chemistry*, D. Reidel, Dordrecht, 1984.
36. Bush, B.L. and Nachbar, R.B., Jr., *J. Comput. Aid. Mol. Des.*, 7 (1993) 587.
37. Waller, C.L., Oprea, T.L., Giolliti, A. and Marshall, G.R., *J. Med. Chem.*, 36 (1993) 4152.
38. Cramer, R.D., III, Bunce, J.D. and Patterson, D.E., *Quant. Struct. Act. Relat.*, 7 (1988) 18.
39. Makhija, M.T. and Kulkarni, V.M., *Bioorg. Med. Chem.*, 2001 (submitted)