

3D-QSAR CoMFA/CoMSIA studies on Urokinase plasminogen activator (uPA) inhibitors: a strategic design in novel anticancer agents

B. A. Bhongade and A. K. Gadad*

Department of Medicinal Chemistry, College of Pharmacy, J. N. Medical College, Belgaum 590010, Karnataka, India

Received 22 December 2003; revised 16 February 2004; accepted 17 February 2004

Available online 1 April 2004

Abstract—Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) was performed on a series of indole/benzoimidazole-5-carboxamides as urokinase plasminogen activator (uPA) inhibitors. The ligand molecular superimposition on template structure was performed by atom/shape-based RMS fit, multfit, and RMSD fit methods. The removal of two outliers from the initial training set of 30 molecules improved the predictivity of the models. The statistically significant model was established from 28 molecules, which were validated by evaluation of test set of nine compounds. The atom-based RMS alignment yielded best predictive CoMFA model ($r_{cv}^2 = 0.611$, $r_{cnv}^2 = 0.778$, F value = 43.825, $r_{bs}^2 = 0.842$, $r_{pred}^2 = 0.616$ with two components) while the CoMSIA model yielded ($r_{cv}^2 = 0.499$, $r_{cnv}^2 = 0.976$, F value = 96.36, $r_{bs}^2 = 0.993$, $r_{pred}^2 = 0.694$ with eight components). The contour maps obtained from 3D-QSAR studies were appraised for the activity trends of the molecules analyzed. The results indicate that the steric, electrostatic, and hydrogen bond donor/acceptor substituents play significant role in uPA activity and selectivity of these compounds. The data generated from the present study can be used as putative pharmacophore in the design of novel, potent, and selective urokinase plasminogen activator inhibitors as cancer therapeutics.

© 2004 Elsevier Ltd. All rights reserved.

1. Introduction

Proteolytic enzymes and their inhibitors play a significant role in various invasive diseases such as the cancer growth and metastasis.¹ Cancer cell invasion, spread, and growth of tumor metastasis is one of the primary cause of mortality and morbidity in malignancy. This invasion is allied with the degradation of basement membrane and other extracellular protein structures implicated in the regulation of extracellular proteolytic events, for instance tissue remodeling, and cell migration.² Tumor cell invasion is a complex, multistep process that involves the controlled degradation of extracellular matrix by tumor cell-associated proteases.³

Urokinase-type plasminogen activator (uPA), a trypsin-like serine protease is strongly associated with tumor

cells, flaunts a vital role in several biological processes including tissue remodeling, cell migration, and matrix degradation through a cascade of mechanisms involving activation of zymogen, plasminogen, and the metalloproteases.^{4–8} uPA cleaves the inactive proenzyme, plasminogen uniquely at Arg⁵⁷⁰–Ser⁵⁷¹ bond generating a broad spectrum serine protease, plasmin, which directly or indirectly impinges the extracellular proteolytic events of many normal and pathological processes like degradation of multiple elements of extracellular matrix, cancer cell invasion etc.^{9–11} Because of its involvement in tumor metastasis and invasion, uPA has emerged as a drug target for development of therapeutics for various types of cancers.^{4–12} This has resulted in an immense clinical interest in developing potent and orally bioavailable inhibitors of uPA that can serve as therapeutic agents in the treatment of cancer. The small molecule inhibitors of uPA has shown to inhibit tumor metastasis and slows cancer growth.^{13–23} However, all of the reported inhibitors of uPA contain an amidino, guanidino, amino group and the compounds designed from templates containing these positively charged moieties could be a liability in the search of novel therapeutic agents.

Keywords: 3D-QSAR; CoMFA; CoMSIA; Indole/benzoimidazole-5-carboxamides; Urokinase plasminogen activator inhibitors; Anticancer agents.

* Corresponding author. Tel.: +91-831-2471399; fax: +91-831-24723-87; e-mail: akgadad@rediffmail.com

Quantitative structure–activity relationship (QSAR) enables the investigators to establish a reliable quantitative structure–activity and structure–property relationships to derive an *in silico* QSAR models to predict the activity of novel molecules prior to their synthesis. The overall process of QSAR model development can be divided into three stages namely; the data preparation, data analysis, and model validation, representing a standard practice of any QSAR modeling. 3D-QSAR methodology have been successfully used to generate models for various chemotherapeutic agents.^{24,25}

We have carried out 3D-QSAR studies employing comparative molecular field analysis²⁶ (CoMFA) and comparative molecular similarity indices analysis²⁷ (CoMSIA) techniques in order to study and deduce a correlation between structure and biological activity of indole/benzoimidazole-5-carboxamidines as uPA inhibitors.^{18,19}

An advantage of CoMFA is its ability to predict the biological activity of molecules and represent the relationship between steric and electrostatic properties and biological activity in the form of contour maps. The CoMSIA technique is of particular interest, as it involves a lipophilic field and two hydrogen bond fields (donor and acceptor) and is less alignment sensitive than CoMFA.

In the present study, we have developed 3D-QSAR CoMFA/CoMSIA models for the series of indole/benzoimidazole-5-carboxamidines and the contour maps derived revealed the significance of steric, electrostatic, hydrogen bond donor, and acceptor fields. The structural variations in the molecular fields at particular regions in the space were studied and 3D-QSAR models generated gives an insight in the design of potent uPA inhibitors.

2. Results

CoMFA and CoMSIA techniques were used to derive 3D-QSAR models for indole/benzoimidazole-5-carboxamidines, as uPA inhibitors. The *in-vitro* inhibitory activity pK_i (μM) was used as dependent variable.

The low energy conformer obtained from systematic search routine was used in the study. All the molecules were aligned employing atom and shape-based RMS fitting, flexible multifit, and RMSD-based database fitting techniques. Using the training set molecules 3D-QSAR models were generated and validated with test set comprising of nine molecules (Table 1). The external predictions were used to select the best model.

Preliminary studies performed using 30 molecules in training set reveal the significance of the CoMFA parameters on final results. PLS analysis was performed with varying σ min. values to set an optimum column filtering value (σ min.). The comparison of F values indicate that fairly good variations been obtained using min. of 1.0 Kcal/mol, therefore a σ min. of 1.0 Kcal/mol

was set in all proceeding calculations as threshold column filtering value in PLS analysis. In order to study the importance of individual CoMFA fields, PLS analysis was performed and the cross-validated r^2 values of only steric fields were found to be higher than that of only electrostatic field and both steric and electrostatic fields analysis, suggesting higher contributions of steric parameters in these series of compounds. The results of PLS analyses are reported in Tables 2–4.

Analysis A shows CoMFA results (Table 2) obtained from four different alignments using 30 molecules in training set. The atom-based RMS alignment I showed cross-validated r^2 0.420 with first two components, non-cross-validated r^2 0.666, F value 26.978, bootstrapped r^2 0.758, predictive r^2 0.600, the steric and electrostatic contributions were 62.8% and 37.2%, respectively.

The CoMFA model generated from shape-based RMS alignment II (Table 2) showed cross-validated r^2 of 0.360 with first two components, non-cross-validated r^2 0.618, F value 21.847, bootstrapped r^2 0.694 and predictive r^2 0.611 with 57.1% steric and 42.9% electrostatic contributions.

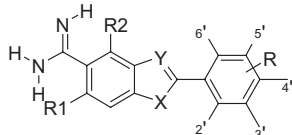
The flexible multifit alignment III (Table 2) yielded cross-validated r^2 of 0.408 with first two components, non-cross-validated r^2 0.645, F value 24.592, bootstrapped r^2 0.750 and predictive r^2 0.613. The steric and electrostatic contributions were 55.1% and 44.9%, respectively.

The RMSD alignment IV (Table 2) yielded cross-validated r^2 of 0.395 with first two components, non-cross-validated r^2 0.640, F value 24.042, bootstrapped r^2 0.746 and predictive r^2 0.603. The steric and electrostatic contributions were 54.3% and 45.7%, respectively.

All the CoMFA models obtained from analysis A demonstrated moderate internal and external predictivity. Thus, in order to increase the predictive power of the derived models, further experiments were performed. Based on the results of QSAR studies from PLS analyses, the two molecules (compounds **29** and **30**) of the training set with high residual values were omitted.

Analysis B (Table 3) shows the CoMFA results obtained from four different alignments using the training set involving 28 molecules and the models derived showed better confidence level in statistical significance. Atom-based RMS aligned model I showed improved cross-validated r^2 value of 0.420–0.611 with the first two components, non-cross-validated r^2 0.666–0.778, F value 26.978–43.825, bootstrapped r^2 0.758–0.842 and a predictive r^2 from 0.600 to 0.616. The steric and electrostatic contributions were found to be 55% and 45%, respectively. The same trend was also observed for rest of the alignments (Table 3) except for shape-based RMS alignment II, in which predictive r^2 decreased from 0.611 to 0.589, with an improved internal predictivity. This model displayed an improved cross-validated r^2 of 0.360 to 0.579 with the first two components, non-cross-validated r^2 from 0.618 to 0.772, F value from 21.847 to

Table 1. Structures and biological activities of training and test set of molecules



Compound	X	Y	R						R1	R2	Biological activity ^a		
			2'	3'	4'	5'	6'	Actual			Calculated		
											CoMFA ^b	CoMSIA	
1 ^c	N	NH	OH	H	H	H	H	H	H	H	-0.740	-0.416	-0.538
2	N	NH	OH	H	H	F	H	H	H	H	-0.447	-0.177	-0.426
3	N	NH	OH	H	H	Br	H	H	H	H	-0.568	-0.340	-0.604
4	N	NH	OH	H	H	Me	H	H	H	H	-0.870	-0.321	-0.751
5	N	NH	OH	H	H	OMe	H	H	H	H	0.556	-0.365	-0.711
6	N	NH	OH	NO ₂	H	H	H	H	H	H	0.045	0.185	0.044
7	N	NH	OH	F	H	H	H	H	H	H	0.259	-0.294	-0.053
8	N	NH	OH	Br	H	H	H	H	H	H	0.250	0.066	0.796
9	N	NH	OH	OMe	H	H	H	H	H	H	-0.556	-0.295	-0.543
10	N	NH	OH	Ph	H	H	H	H	H	H	0.390	0.596	0.590
11	N	NH	OH	H	Me	H	H	H	H	H	-0.556	-0.400	-0.861
12	N	NH	OH	H	H	H	OH	H	H	H	-0.278	-0.341	-0.189
13	N	NH	OH	-naphthyl-		H	H	H	H	H	-0.591	-0.071	-0.635
14	N	NH	OH	Br	H	Me	H	H	H	H	0.552	0.175	0.597
15	CH	NH	OH	H	H	H	H	H	H	H	-0.380	-0.041	-0.094
16	CH	NH	OH	Br	H	Me	H	H	H	H	1.250	0.626	1.234
17	CH	NH	OH	Br	H	H	H	H	H	H	1.450	0.503	1.405
18	CH	NH	OH	Br	H	NO ₂	H	H	H	H	0.552	0.642	0.480
19	CH	NH	OH	Ph	H	Me	H	H	H	H	1.420	1.693	1.431
20	CH	NH	OH	Ph	H	NO ₂	H	H	H	H	1.602	1.186	1.679
21	CH	NH	H	Ph	H	H	H	H	H	H	-1.113	-0.700	-1.163
22	CH	NH	OMe	Ph	H	H	H	H	H	H	-0.500	-0.257	-0.463
23	NMe	N	OH	Ph	H	H	H	H	H	H	-1.041	-1.155	-1.091
24	CH	O	OH	Ph	H	H	H	H	H	H	0.259	-0.441	0.288
25	CH	NH	OH	Ph	H	H	H	F	H	H	1.698	1.327	1.765
26	CH	NH	OH	Ph	H	H	H	OMe	H	H	0.440	1.292	0.588
27	CH	NH	OH	Ph	H	H	H	H	H	H	2.096	1.561	1.766
28	CH	NH	OH	Ph	H	H	H	Cl	H	H	2.045	1.567	1.920
29	CH	NH	OH	Ph	H	H	H	OH	H	H	-0.300	1.321	1.426
30	CH	NH	OH	Ph	H	H	H	H	Cl	H	-0.477	1.367	1.753
31 ^d	N	NH	OH	H	H	NO ₂	H	H	H	H	-0.518	-0.177	-1.270
32	N	NH	OH	H	NEt ₂	H	H	H	H	H	-1.113	-0.519	-1.710
33	N	NH	OH	Ph	H	Cl	H	H	H	H	0.301	0.684	0.563
34	CH	NH	OH	Br	H	Br	H	H	H	H	1.000	0.562	1.397
35	CH	NH	OH	Ph	H	Br	H	H	H	H	1.360	1.660	1.626
36	CH	NH	OH	Ph	H	Cl	H	H	H	H	1.250	1.670	1.706
37	CH	NH	OH	Ph	H	H	H	Me	H	H	1.000	1.512	1.342
38	N	NH	H	H	H	H	H	H	H	H	-0.903	-0.351	-0.933
39	N	NH	OH	Me	H	H	H	H	H	H	-1.740	0.089	0.188

^a Biological activity expressed as log 1/*K_i* against human uPA enzyme in μM.^b Calculated activity from atom-based RMS alignment I.^c Training set 1–30.^d Test set 31–39.

42.429 and bootstrapped r^2 from 0.694 to 0.849 with 58.2% and 41.8% steric and electrostatic contributions, respectively.

The flexible multifit alignment III (Table 3) yielded cross-validated r^2 of 0.615 with first two components, non-cross-validated r^2 of 0.780, F value of 44.302, bootstrapped r^2 of 0.850 and predictive r^2 of 0.555. The steric and electrostatic contributions were 56.5% and 43.5%, respectively.

The RMSD alignment IV (Table 3) yielded cross-validated r^2 of 0.594 with first two components, non-cross-validated r^2 of 0.770, F value of 41.804, bootstrapped r^2 of 0.851 and predictive r^2 of 0.547. The steric and electrostatic contributions were found 56.0% and 44.0%, respectively.

Thus, all the CoMFA models derived from analysis B showed higher contributions of steric parameters toward the activity of these compounds similar to that of

Table 2. Summary of CoMFA results (analysis A)

Alignment	I ^a	II ^b	III ^c	IV ^d
r^2_{cv} ^e	0.420	0.360	0.408	0.395
Components	2	2	2	2
SEP ^f	0.751	0.798	0.759	0.768
r^2_{ncv} ^g	0.666	0.618	0.645	0.640
SEE ^h	0.570	0.610	0.588	0.592
<i>F</i> value	26.978	21.847	24.592	24.042
P $r^2_{=0}$	0	0	0	0
Contrib. Steric	62.8	57.1	55.1	54.3
Contrib. Elect.	37.2	42.9	44.9	45.7
r^2_{pred} ⁱ	0.600	0.611	0.613	0.603
r^2_{bs} ^j	0.758	0.694	0.750	0.746
Standard deviation ^j	0.068	0.078	0.083	0.070

^a Alignment by atom-based RMS fit.^b Alignment by shape-based RMS fit.^c Alignment by Multifit.^d Alignment by RMSD database.^e Cross-validated r^2 by leave-one-out method.^f Standard error of prediction.^g Non-cross-validated r^2 .^h Standard error of estimate.ⁱ Predictive r^2 .^j From 100 bootstrapping runs.**Table 3.** Summary of CoMFA results (analysis B)

Alignment	I ^a	II ^b	III ^c	IV ^d
r^2_{cv} ^e	0.611	0.579	0.615	0.594
Components	2	2	2	2
SEP ^f	0.630	0.656	0.627	0.664
r^2_{ncv} ^g	0.778	0.772	0.780	0.770
SEE ^h	0.478	0.482	0.474	0.485
<i>F</i> value	43.825	42.429	44.302	41.804
P $r^2_{=0}$	0	0	0	0
Contrib. Steric	55.0	58.2	56.5	56.0
Contrib. Elect.	45.0	41.8	43.5	44.0
r^2_{pred} ⁱ	0.616	0.589	0.555	0.547
r^2_{bs} ^j	0.842	0.849	0.850	0.851
Standard deviation ^j	0.053	0.048	0.043	0.053

^a Alignment by atom-based RMS fit.^b Alignment by shape-based RMS fit.^c Alignment by Multifit.^d Alignment by RMSD database.^e Cross-validated r^2 by leave-one-out method.^f Standard error of prediction.^g Non-cross-validated r^2 .^h Standard error of estimate.ⁱ Predictive r^2 .^j From 100 bootstrapping runs.

analysis A. Based on the predictive ability of the four CoMFA models from analysis B (Table 3), the model generated with atom-based RMS alignment I carrying good predictive r^2 of 0.616 was selected for further analysis and all the CoMFA contours were generated using this model. The graphs of actual versus fitted and predicted activities for the training and test set of molecules are depicted in Figures 2 and 3, respectively. The field values generated at each grid point were calculated as the scalar product of the associated QSAR coefficient and the standard deviation of all values in the corresponding column of the data table (STDDEV*COEFF), plotted as the percentage contributions to QSAR

equation. The CoMFA steric and electrostatic contour maps developed using the atom-based RMS alignment I analysis are shown in Figures 4 and 5, respectively.

The CoMFA steric map (Fig. 4) encompasses yellow contours (80% contribution) corresponding to the regions in space where the steric bulk envisages the decrease in activity and the yellow polyhedron bordering the 3'-phenyl ring suggests that the steric substitutions at phenyl ring are not favorable. Conversely, the green regions (20% contribution) embedded in the 3'-phenyl ring and in the vicinity of 3-H of compound **27** reveals that, an increase in activity is anticipated due to increased steric bulk.

CoMFA electrostatic map (Fig. 5), displays red contours at the vicinity of 6-H and space between 2'-OH and 3'-phenyl ring where the partial negative charge is associated with increased activity (80% contribution). Blue contours (20% contribution) observed at the upper portion of 2-phenyl ring and at the vicinity of 3'-phenyl ring of compound **27** indicates areas within the lattice where the electropositive properties of molecules describe an increase in activity.

In addition to the steric and electrostatic fields, CoMSIA also defines the lipophilic, hydrogen bond donor, and acceptor fields that are generally not accessible with standard CoMFA. The atom based RMS alignment I used in the CoMFA studies served as an alignment for CoMSIA and the results of the study are summarized in Table 4.

The CoMSIA model with the combination of all fields yielded a cross-validated r^2 0.499 with eight components, non-cross-validated r^2 0.976, *F* value 96.369, bootstrapped r^2 0.993 with highest predictive r^2 0.694 and the steric, electrostatic, hydrogen bond donor/acceptor, and hydrophobic field contributions of this model were 11.2%, 25.6%, 33.5%, 13.1%, and 16.6%, respectively.

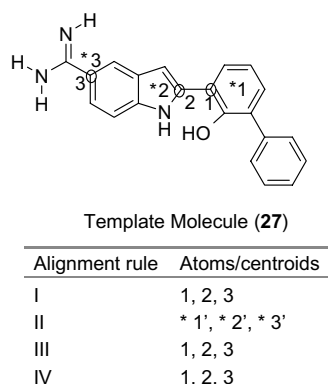
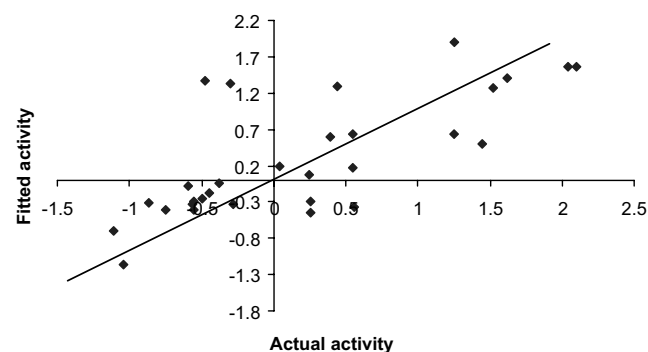
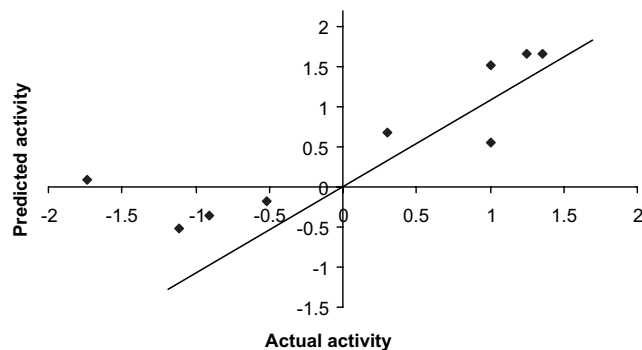
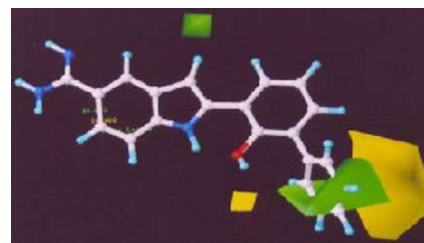
CoMSIA model with the combination of steric, electrostatic, and hydrogen bond donor fields yielded a highest cross-validated r^2 0.628 with first five components, non-cross-validated r^2 0.952, *F* value 86.680, bootstrapped r^2 0.976 and predictive r^2 0.419, 23.8% steric, 47.3% electrostatic, and 28.8% hydrogen bond donor field contributions.

The combinations of hydrophobic, hydrogen bond donor, and acceptor fields yielded CoMSIA model with a highest predictive r^2 0.792. This model also showed cross-validated r^2 of 0.410 with nine components, non-cross-validated r^2 0.961, *F* value 49.13, bootstrapped r^2 0.987, 23.8% steric, 47.3% electrostatic, and 28.8% hydrogen bond donor field contributions. The models generated by various combinations of CoMSIA fields (Table 4) showed statistically significant moderate to high internal and external predictions in which the electrostatic and hydrogen bond donor fields made the higher contributions to the CoMSIA QSAR models. The graphs of actual versus fitted and predicted

Table 4. Summary of CoMSIA results

	S ^a	E ^b	D ^c	A ^d	H ^e	S+E	S+E+D	S+E+A	S+E+H	D+A	D+A +S	D+A +E	D+A +H	S+E+ D+A	ALL
r_{cv}^{2f}	0.483	0.612	0.438	0.344	0.368	0.584	0.628	0.585	0.430	0.488	0.567	0.576	0.410	0.601	0.499
N_c^g	2	2	1	5	7	2	5	2	2	3	5	5	9	5	8
SEP ^h	0.727	0.840	0.743	0.873	0.886	0.658	0.658	0.652	0.763	0.738	0.710	0.702	0.915	0.681	0.821
r_{ncv}^{2i}	0.606	0.756	0.605	0.722	0.941	0.717	0.952	0.742	0.724	0.800	0.902	0.924	0.961	0.934	0.976
SEE ^j	0.635	0.500	0.623	0.568	0.275	0.538	0.237	0.513	0.531	0.461	0.337	0.298	0.236	0.277	0.180
F value	19.20	38.65	39.88	11.42	45.40	31.70	86.68	35.97	32.79	32.06	40.67	53.26	49.13	62.12	96.36
$Pr^2 = 0$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r_{pred}^{2k}	—	—	—	—	—	0.536	0.419	0.637	0.559	0.717	0.579	0.600	0.792	0.567	0.694
r_{bs}^{2l}	0.670	0.813	0.638	0.798	0.974	0.776	0.976	0.792	0.763	0.842	0.938	0.959	0.987	0.962	0.993
Std.	0.093	0.050	0.107	0.078	0.015	0.071	0.011	0.060	0.082	0.056	0.026	0.018	0.009	0.013	0.005
Dev. ^l															

ALL = S+E+D+A+H.

^a Steric.^b Electrostatic.^c Hydrogen bond donor.^d Hydrogen bond acceptor.^e Hydrophobic.^f Cross-validated r^2 by leave-one-out method.^g Number of components.^h Standard error of prediction.ⁱ Non-cross-validated r^2 .^j Standard error of estimate.^k Predictive r^2 .^l From 100 bootstrapping runs.**Figure 1.** Atoms and centroids used in superimposum of the molecules.**Figure 2.** A graph of actual versus fitted activities of the training set molecules from atom-based RMS alignment I (analysis B).**Figure 3.** A graph of actual versus predicted activities of the test set molecules from atom-based RMS alignment I (analysis B).**Figure 4.** CoMFA steric STDDEV*COEFF contour plots from the atom-based RMS alignment I (analysis B). Sterically favored areas (contribution level of 80%) are represented by green polyhedra. Sterically disfavored areas (contribution level of 20%) are represented by yellow polyhedra. The active compound 27 in ball and stick is shown.

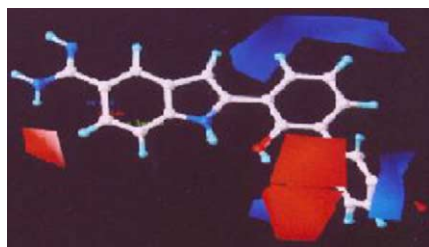


Figure 5. CoMFA electrostatic STDDEV*COEFF contour plots from the atom-based RMS alignment I (analysis B). Positive charged favored areas (contribution level of 80%) are represented by blue polyhedra. Negatively charged favored areas (contribution level of 20%) are represented by red polyhedra. The active compound **27** in ball and stick is shown.

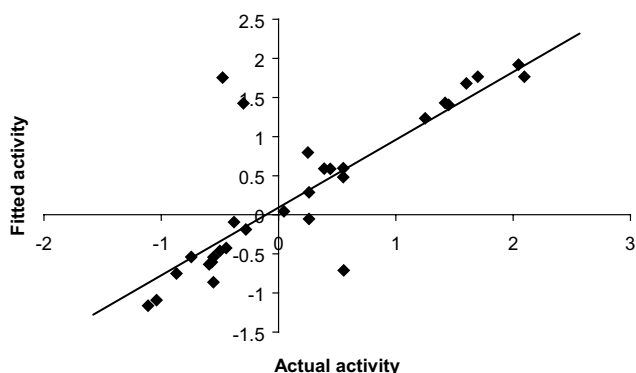


Figure 6. A graph of actual versus fitted activities of the training set molecules from CoMSIA analysis with steric, electrostatic, and hydrogen bond donor fields.

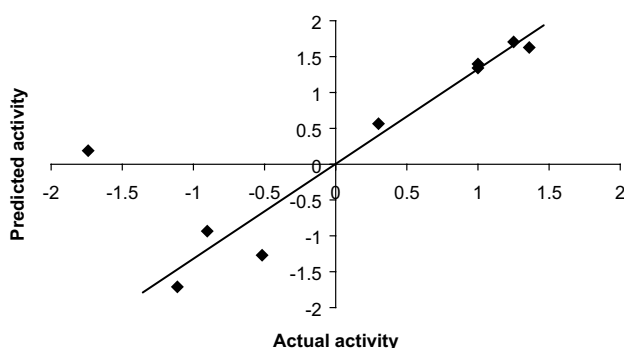


Figure 7. A graph of actual versus predicted activities of the test set molecules from CoMSIA analysis with steric, electrostatic, and hydrogen bond donor fields.

activities for the training and test set molecules from all CoMSIA fields are shown in Figures 6 and 7, respectively. The CoMSIA steric and electrostatic contours (not shown) were placed almost similar to that of the CoMFA model.

The hydrogen bond donor and acceptor contour maps of CoMSIA (STDDEV*COEFF) are displayed in Figures 8 and 9, respectively. The CoMSIA hydrogen bond donor contour map (Fig. 8) indicate cyan contours

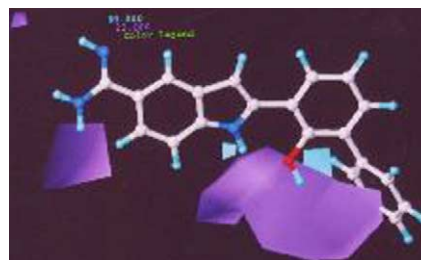


Figure 8. CoMSIA hydrogen bond donor fields.

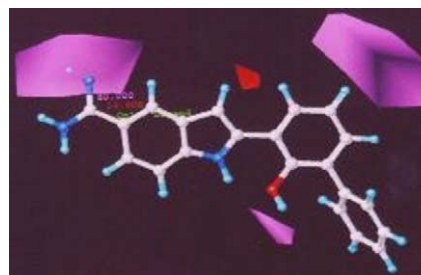


Figure 9. CoMSIA hydrogen bond acceptor fields.

(80% contribution) near 1-NH and 2'-CH of 3'-phenyl group as the favorable region while the prominent purple contours (20% contribution) at lower region of 3'-phenyl group and near 6-H, 5-amidino group of compound **27** as the disfavored regions for hydrogen bond donor substituents. The hydrogen bond acceptor contour map (Fig. 9) connotes magenta contours (80% contribution) in the proximity of 5-amidino, 2'-hydroxyl and 5'-H indicates favorable regions and the disfavored regions are shown by the red contours (20% contribution) near 3-H of compound **27**.

3. Discussion

In the present 3D-QSAR studies, CoMFA, and CoMSIA methodologies were employed to a series of indole/benzimidazole-5-carboxamidines as uPA inhibitors. The most critical and important part of the QSAR model development is the model validation, where the internal predictive power of model and its ability to reproduce biological activities of untested compounds is to be established. This essentially depends on the orientation of ligands and selection of training/test set molecules. The alignment defines the putative pharmacophore for the series of ligands and predictive power of the QSAR models reveals the significance of alignment in 3D-QSAR model development. The ligand molecules were aligned onto a template structure (compound **27**) employing four different alignment rules (Fig. 1) and all the alignments exhibited statistically significant correlative models with an average to good predictivity supporting our choice of atoms for superimposition. The variations in predictivity from the different alignment rules employed in the present study may be due to the

rigidity of ligands in test set and a slight change in their orientations leads to the placement of functional groups in unfavorable regions supporting the exact superimposition of ligand molecules on the template structure essential for good predictions in 3D-QSAR/CoMFA studies. The CoMFA models derived using 30 molecules in training set yielded poor internal predictivity (Table 2), while an improved internal and external predictivity was observed with 28 molecules in training set (Table 3). The two outliers (compounds **29** and **30**) with highest residues from the non-cross-validated PLS QSAR analysis were identified and their poor prediction may be due to the composition of training set molecules. The 6-substituted compounds of the training set possessing small electronegative groups caused an increase in the activity (compounds **25** and **28**), while the compounds **29** and **30** showed decreased activity for the reason that compound **29** was the only compound in these series with a lipophilic hydroxyl substitution at the sixth position and the compound **30** as the only 4-substituted (chloro) compound amongst the training set molecules. Similarly, the poor prediction for the test set compound **39** with 3'-methyl group is anticipated due to the presence of alkyl group at 3'-position, which favors the hydrophobic substituents.

Atom-based RMS I aligned CoMFA model (Table 3) with good internal and maximum external predictivity was used in the analysis of CoMFA contours (Figs. 4 and 5). Steric substituents at 3' position of 2-phenyl ring favors the activity (compounds **10** and **27**) while sterically unfavorable contours surrounding 3'-phenyl rings suggested that the substructure fragments with less or high steric bulk in this area reduces the activity (compounds **11** and **13**). In electrostatic contour plots (Fig. 5), red contours at the vicinity of 6-H and space between 2'-OH and 3'-phenyl ring suggests that increased activity is anticipated by moderate electronegative substituents at 6-position (compounds **25** and **28**) and 3'-position of 2-phenyl ring (compounds **7**, **8**, and **17**), where as the blue contour indicate low electron density substituents at 5', 6'positions of 2-phenyl ring favoring the activity (compound **5**). These regions include the partial positive charges associated with hydrogen atoms bound to carbon, which can be correlated with lipophilic interactions.

The CoMSIA QSAR studies gives an additional structural insight in the study of the probable binding sites of the ligand-receptor. The hydrogen bond donor favorable cyan contour (Fig. 8) in the vicinity of 1-NH of indole/benzoimidazole shows the significance of proton in the hydrogen bond formation with the receptor surface. The replacement of 1-NH with N (compound **23**) and O (compound **24**) resulted in reduced activity. The hydrogen bond donor disfavorable purple contours (Fig. 8) appears at lower region of 3'-phenyl group and near 6-H, 5-amidino group of compound **27**.

The favorable magenta contours for the hydrogen bond acceptor substituents (Fig. 9) highlight the significance of 5-amidino, 2'-hydroxyl groups in the activity and selectivity of these compounds. The replacement of 2'-

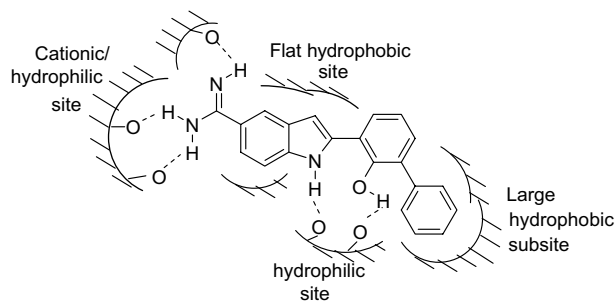


Figure 10. Proposed hypothetical model for uPA binding site.

hydroxyl group with H (compound **21**), OMe (compound **22**) resulted in decreased activity. The results of CoMSIA studies (Table 4), demonstrate the significance of the positively charged 2'-OH, 5-amidino groups in binding the receptor.

Based on the present 3D-QSAR studies, hypothetical binding model of these ligand molecules with uPA can be proposed (Fig. 10). In the cationic/hydrophilic site, ligands may form hydrogen bonds with the active sites of the receptor. In the flat hydrophobic linker region, the aromatic rings may have π - π interactions with the receptor where the absolute planarity in the ligand structure is essential. This is the most important region of ligand molecules, which can be explored to design potent uPA inhibitors. The large hydrophobic region/binding site may play a significant role in the selectivity of ligands over the counterparts of the uPA and a slight variation in the orientation of phenyl ring results in decreased activity as seen in 3D-QSAR studies.

The present results of 3D-QSAR studies and the proposed hypothetical ligand binding model are in accordance with the interactions between the basic group and Asp¹⁸⁹ in S1 subsite and 2'-hydroxyl group with catalytic residues as essential components for binding with uPA.^{18,19}

4. Conclusion

As the crystallographic data for these ligand-uPA complex was not available in protein data bank, hence a receptor independent 3D-QSAR has been established for indole/benzoimidazole-5-carboxamides as uPA inhibitors employing the most widely used techniques CoMFA and CoMSIA. All the molecular modeling and 3D-QSAR studies were performed with the standard protocol using SYBYL 6.7 software. Present studies signify the importance of ligand orientation and selection of the training set molecules in the development of statistically significant QSAR models. Overall, the CoMSIA method provided better statistical models than CoMFA, which implies the significance of hydrogen bond donor/acceptor fields in the selectivity and activity of these ligands in addition to the steric and electrostatic fields. The statistical significance and robustness of

generated 3D-QSAR models were confirmed using an external set of molecules.

The 3D-QSAR models generated in the present study are consistent with the binding site^{18,19} and can be used as putative pharmacophore. The structural requirements identified in the present study can be utilized strategically in the design of novel, potent, and selective urokinase plasminogen activator inhibitors as anticancer agents.

5. Methods

5.1. Biological data

Series of indole/benzimidazole-5-carboxamidines have been reported as selective uPA inhibitors and tested against human trypsin-like serine protease, uPA.

Ethylenediaminetetraacetic acid (EDTA) has been used to remove zinc ion, this eliminates zinc mediated binding effects in the assay for some of these scaffolds. The assay was buffered at pH 7.4 in order to mimic physiological pH (for detail method see Refs. 18 and 19). Apparent inhibition constants were converted to K_i values, the negative logarithm of measured K_i (μM) against uPA enzyme as $\text{p}K_i$ was used in 3D-QSAR ($\text{p}K_i = \log 1/K_i$). Rationally selected training and test set provide a means for the development of reliable validated QSAR models.²⁸ Table 1 defines structures and biological activity of training set (1–30 molecules) and test set (31–39 molecules). The robustness and predictive ability of models were evaluated by selecting wide range of biological activity with chemical classes similar to the training set.

5.2. Molecular modeling

All molecular modeling and 3D-QSAR studies were performed using SYBYL 6.7²⁹ with standard TRIPOS force field³⁰ on Silicon Graphics Indy O2 workstation. The crystallographic data for these ligand–uPA complex was not available in protein data bank hence, all the molecules were constructed using standard geometric and bond lengths with SYBYL, the initial optimization was carried using standard TRIPOS force field employing the Gasteiger Marsili charges, the constraints were removed and repeated minimization was performed using step descent and conjugated gradient method till the RMSD of 0.001 kcal/mol was achieved. Due to lack of structural data supporting a specific ‘active’ conformation, we have assumed that a compound is active in an energetically minimized conformation. Systematic search routine was used in the conformational analysis and all rotatable bonds were searched in 10° increments from 0° to 360°. Conformational energies were computed with electrostatic term and the low energy conformers were selected for superimpositions. Further optimization was performed

using MOPAC with AM1 Hamiltonian³¹ and MOPAC charges derived were used in subsequent analysis.

5.3. Alignment rules

The molecular conformation and orientation is one of the most sensitive input areas in 3D-QSAR studies. In present study, superimposition of the molecules was carried out by different approaches using compound **27** (Table 1) as template structure.

- Alignment I In this alignment, atoms of the molecules were used for RMS fitting on corresponding atoms of the template.
- Alignment II Here, the centroids rather than exact superimposition of atoms of the rings were used for RMS fitting to the template.
- Alignment III The alignment of molecules was carried out by flexible fitting (multifit) of the atoms of molecules to the template.
- Alignment IV Each analog was aligned to the template by rotation and translation so as to minimize RMSD between atoms in the template and the corresponding atoms in the analog using the DATABASE ALIGN command in SYBYL. The atoms/centroids used for alignments are defined in Figure 1. These alignments were subsequently used in CoMFA/CoMSIA probe interaction energy calculations.

5.4. CoMFA and CoMSIA studies

The steric and electrostatic field energies were calculated using Sp^3 carbon probe atom with van der Waals radius of 1.52 Å and +1 charge. These energies were truncated to ± 30 kcal/mol and the electrostatic contributions were ignored at lattice interactions with maximum steric interactions. The CoMFA fields generated automatically were scaled by CoMFA-STD method in SYBYL.

Similarity indices descriptors were derived with the same lattice box, which were used in CoMFA calculations. Five CoMSIA similarity index fields available within SYBYL (steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor) were evaluated using the probe atom. Gaussian-type distance dependence was between the grid point and each atom of molecule and a default value of 0.3 was used as attenuation factor.

5.5. Partial least square (PLS) analysis

The CoMFA/CoMSIA descriptors served as independent variables and $\text{p}K_i$ values as dependent variables in PLS regression analysis³² in deducing the 3D-QSAR models. Normally cross-validation is used to check the predictivity of the derived model. The result of analyses corresponds to a regression equation with thousands of

coefficients. The predictive values of models were evaluated using leave-one-out (LOO) cross-validation method. The optimum number of components (Nc) used to derive the nonvalidated model was defined as the number of components leading to the highest cross-validated r^2 and lowest standard error of prediction (SEP). To obtain the statistical confidence limit in the analyses, PLS analysis using 100 bootstrap groups with optimum numbers of components was performed.

5.6. Predictive r^2 values

The predictive ability of each analysis was determined from a test set of nine compounds that were not included in the training set. These molecules were aligned and their activities were predicted by each PLS analysis. The predictive r^2 (r^2_{pred}) value defined as

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

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

Acknowledgements

Authors express their thanks to Dr. F. V. Manvi, Principal, Prof. A. D. Taranalli, Vice-Principal, College of Pharmacy, Belgaum, India, for providing necessary facilities. B.A.B. is thankful to HRDG, Council for Scientific and Industrial Research, New Delhi for the award of Senior Research Fellowship [8/470(1)/2001–EMR-I]. Authors also acknowledge All India Council for Technical Education, New Delhi for the financial support [File No. 8018/RDII/BOR/TAP (397)/99-2000].

References and notes

- Reuning, U.; Magdolen, V.; Wilhelm, O.; Fischer, K.; Lutz, U.; Graeff, H.; Schmitt, M. *Int. J. Oncol.* **1998**, *13*, 893.
- Strickland, S.; Reich, E.; Sherman, M. T. *Cell* **1996**, *9*, 231.
- Mangatti, P.; Rafkin, D. B. *Physiol. Rev.* **1993**, *73*, 161.
- Andreassen, P. A.; Kjoller, L.; Christensen, L.; Duffy, M. J. *Int. J. Cancer* **1997**, *72*, 1.
- Magill, C.; Katz, B. A.; Mackman, R. L. *Emerg. Ther. Targets* **1999**, *3*, 109.
- Rabbani, S. A.; Xing, R. H. *Int. J. Oncol.* **1998**, *12*, 911.
- Blasi, F. *Bioassays* **1993**, *15*, 105.
- Ellis, V.; Behrendt, N.; Dano, K. *J. Biol. Chem.* **1991**, *266*, 12752.
- Dano, K.; Andreassen, P. A.; Grondahl, H. J.; Kristensen, P.; Neilsen, L. S.; Skriver, L. *Adv. Cancer Res.* **1985**, *44*, 139.
- Testa, J. E.; Quigley, J. P. *Cancer Metast. Rev.* **1998**, *9*, 353.
- Mangatti, P.; Robbins, E.; Rafkin, D. B. *Cell* **1986**, *47*, 487.
- Meyer, T.; Hart, I. R. *Eur. J. Cancer* **1998**, *34*, 214.
- Vassali, J.-D.; Belin, D. *FEB* **1987**, *214*, 187.
- Yong, H.; Henkin, J.; Kim, K. H.; Greer, J. J. *Med. Chem.* **1990**, *33*, 2956.
- Towle, M. J.; Lee, A.; Maduakor, E. C.; Schwartz, E. C.; Bridges, A. J.; Littlefield, B. A. *Cancer Res.* **1993**, *53*, 2553.
- Sperl, S.; Jacob, U.; dePrada, N. A.; Sturzebecher, J.; Wilhelm, O. G.; Bode, W.; Magdolen, V.; Huber, R.; Moroder, L. *PNAS* **2000**, *97*, 5113.
- Hajduk, P. J.; Boyd, S.; Nettesheim, D.; Nienaber, V.; Severin, J.; Smith, R.; Davidson, D.; Rockway, T.; Fesik, S. W. *J. Med. Chem.* **2000**, *43*, 3862.
- Verner, E.; Katz, B. A.; Spencer, J. R.; Allen, D.; Hataye, J.; Hruzewicz, W.; Hui, N. C.; Kolesnikov, A.; Li, Y.; Loung, C.; Martelli, A.; Kesavan, R.; Rai, R.; She, M.; Shrader, W.; Sprengeler, P. A.; Trapp, S.; Wang, J.; Young, W. B.; Mackman, R. L. *J. Med. Chem.* **2001**, *44*, 2753.
- Mackman, R. L.; Katz, B. A.; Breitenbucher, J. G.; Hui, H. C.; Loung, C.; Liu, L.; Sprengeler, P. A. *J. Med. Chem.* **2001**, *44*, 3856.
- Barber, C. G.; Dickinson, R. P.; Horne, V. A. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 181.
- Barber, C. G.; Dickinson, R. P. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 185.
- Mackman, R. L.; Hui, H. C.; Breitenbucher, J. G.; Katz, B. A.; Loung, C.; Martelli, A.; McGee, D.; Kesavan, R.; Sendzik, M.; Spencer, J. R.; Sprengeler, P. A.; Tario, J.; Verner, E.; Wang, J. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2019.
- Spencer, J. R.; McGee, D.; Allen, D.; Katz, B. A.; Luong, C.; Sendzik, M.; Squires, N.; Mackman, R. L. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2023.
- Gokhale, V. M.; Kulkarni, V. M. *Bioorg. Med. Chem.* **2000**, *8*, 2487.
- Karki, R. G.; Kulkarni, V. M. *Bioorg. Med. Chem.* **2001**, *9*, 3153.
- Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
- Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130.
- Golbraikh, A.; Shen, M.; Xiao, Z.; Xiao, Y.; Lee, K.; Tropsha, A. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 241.
- Sybyl 6.7, Tripos Associates, 1699 South Hanley Road, St. Louis, MO 63144.
- Clark, M.; Cramer, R.-D.; Van, O. N. *J. Comput. Chem.* **1989**, *10*, 982.
- Stewart, J. J. P. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1.
- Wold, S.; Albano, C.; Dunn, W.; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjostrom, M. In *Chemometrics: Mathematics and Statistics in Chemistry*; Kowalski, B., Ed.; Reidel: Dordrecht, The Netherlands, 1987; p 17.