Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 13 (2005) 2773–2782
Bioorganic & Medicinal Chemistry

3D-QSAR CoMFA studies on trypsin-like serine protease inhibitors: a comparative selectivity analysis

Bhoomendra A. Bhongade,[a] Veerappa V. Gouripur[b] and Andanappa K. Gadad[a,b,*]

[a]*Department of Medicinal Chemistry, College of Pharmacy, J. N. Medical College, Belgaum 590 010, Karnataka, India*
[b]*Pharmacy Programme, Faculty of Medical Sciences, The University of The West Indies, St. Augustine, Champs Fluers, Mount Hope, Trinidad, WI*

Received 5 January 2005; revised 14 February 2005; accepted 15 February 2005

Dedicated to Professor Vithal M. Kulkarni

Abstract—A series of indole/benzoimidazole-5-carboxamidines have been reported to inhibit various trypsin-like serine proteases viz. uPA, tPA, factor Xa, thrombin, plasmin, and trypsin, which are involved in various types of pathophysiological conditions such as cancer progression, thrombosis etc. Inhibition of these protease enzymes may serve as therapeutic agents in various types of cancer as well serve as anticoagulant or antithrombotic agents. The dual inhibitory action may result in poor clinical candidates. 3D-QSAR models were generated for indole/benzoimidazole-5-carboxamidines using the CoMFA technique to study their selectivity trends toward various trypsin-like serine proteases. Molecular superimposition was carried out on the template structure using atom-based RMS fit method. The CoMFA models were established from the training set of 25–29 molecules and validated by predicting the activities of seven–eight test set molecules. The CoMFA models generated using steric and electrostatic fields for tPA, fXa, thrombin, plasmin, and trypsin inhibition exhibited better statistical significance than the CoMFA models generated using ClogP as an additional descriptor. Thus, the validated CoMFA models with steric and electrostatic fields were used to generate 3D contour maps, which may provide possible modification of molecules for better selectivity/activity. The present 3D-QSAR studies emphasize the selectivity trends of indole/benzoimidazole-5-carboxamidines, which may be obliging in designing novel selective serine protease inhibitors of therapeutic interest.
© 2005 Elsevier Ltd. All rights reserved.

1. Introduction

Trypsin-like serine protease plays the central role in a wide range of biological processes and the upregulation of the members of this superfamily is therefore an attractive target for the development of potential therapeutic agents. Most of these targets and their anti-targets are characterized depending on the presence of amino acid residues (threonine, serine, or alanine) at the position 190 of S1 site. Within this superfamily of proteolytic enzymes, several potential drug targets viz. the fibrinolytic enzyme, plasmin and its activators, urokinase-type plasminogen activator (uPA), tissue-type plasminogen activator (tPA), the precoagulant enzyme factor Xa (fXa), IXa, VIIa, IIa (thrombin), trypsin, and the anticoagulant activated protein C, etc., have been identified. uPA, trypsin, and tryptase possess serine residue, whereas tPA, thrombin, and fXa posses an alanine residue at position 190.

The fibrinolytic system of mammalian blood comprises of an inactive proenzyme plasminogen, which is converted to an active enzyme plasmin by two immunologically distinct physiological plasminogen activators: tPA and uPA. The function of plasminogen activators in biological processes other than the extracellular activation of plasminogen has been proposed based on differences in the tissue distribution of these two forms of plasminogen activators.[1] The process of plasminogen activation is central to both extracellular matrix remodeling and the cell invasiveness mediated by uPA,[2,3] while fibrinolysis in vascular space is mediated by tPA.[4–6]

Plasmin, a broad spectrum protease degrades most of the protein components of the extracellular matrix

Keywords: 3D-QSAR; CoMFA; Indole/benzoimidazole-5-carboxamidines; Trypsin-like serine protease inhibitors; Anticancer; Anticoagulants; Antithrombotic agents.
* Corresponding author. Tel.: +91 831 2471399/+1 868 662147; fax: +91 831 2472387; e-mail: akgadad@rediffmail.com

0968-0896/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2005.02.027

(ECM) directly by removing the glycoproteins from ECM or by activation of matrix metalloproteinases (MMPs). It converts inactive MMPs to the active form responsible for ECM proteolysis and prevents neutralization of the MMP blocking the metalloproteinase inhibitors.[7,8] Several malignant tumors including human malignant gliomas have shown to produce uPA[9,10] and the cells possess specific receptor for uPA, that is, uPAR provides an inducible, transient and localized cell surface for proteolytic activity. Tumor invasiveness and metastasis can be blocked by direct reduction of uPA activity.[11–13] It is also a prognostic marker in the human breast, lung cancer and gliomas where high levels of uPA are associated with decreased survival.[10,14] Because of its involvement in tumor metastasis and invasion, uPA has emerged as a novel therapeutic target for cancer.[15]

The other protease enzyme human fXa located at the convergence of the intrinsic and extrinsic blood coagulation pathways catalyze the activation of prothrombin to thrombin,[16,17] and flaunts a major role in thrombosis and haemostasis.[18] Factor Xa possesses focal function and upstream location from thrombin in the coagulation cascade. The selective inhibitors of fXa compared to thrombin may be safe and effective in the treatment of thrombotic disorders making fXa an attractive target for anticoagulant drug development.

The structures of various trypsin-like serine protease available as potential therapeutic targets including human α-thrombin,[19] fXa,[20] tPA,[21] and uPA[22] have opened a new doorway in the structure-based design of serine protease inhibitors. The members of the trypsin-like serine protease family possesses a catalytic site with high sequence homology and similar substrate specificity, therefore their inhibitors need to be highly potent, selective among the closely related enzymes with adequate pharmacodynamic properties. The published structures offer the possibility of understanding the specificity over related counterparts/enzymes with high sequence homology and similar substrate specificity.

The 3D-QSAR/CoMFA[23] generates steric and electrostatic contour maps in the 3D space around the molecules where changes in the respective physicochemical properties are predicted to increase or decrease the biological activity. The contour maps may aid in visualizing the comparative region of differences in the related enzymes and also suggest for the flexibility in protein structure, which may not be observed in the X-ray structure analysis.

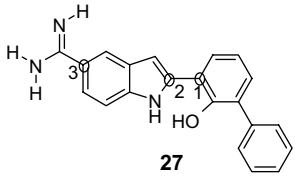
Most of the known small molecule inhibitors of serine proteases bear an amidino/guanidino group as the pharmacophoric group, which may help to produce high binding affinity with the protease enzymes due to its interaction with Asp189 in the S1 specificity pocket.[24] The differences in the S1 subsite for each enzyme can be exploited for specificity.[24–26] The potency and selectivity of the target of interest over the related enzymes is mainly responsible for the interactions of the ligand with the neighboring subsites differing from their contourparts. Thus, achieving selectivity becomes critical and is typically generated through the elaboration of the chemical groups that occupy subsites around the catalytic machinery.

As part of our ongoing research in the development of novel selective uPA inhibitors, we had previously reported the 3D-QSAR CoMFA/CoMSIA models for indole/benzoimidazole-5-carboxamidines as uPA inhibitors.[27] In continuation of the research work, it was thought worthwhile to study the selectivity trends of these inhibitors toward the other trypsin-like serine protease, which may provide additional information in designing the selective uPA inhibitors.

In the present paper, we report the 3D-QSAR/CoMFA models developed for the series of indole/benzoimidazole-5-carboxamidines[28,29] as trypsin-like serine protease inhibitors viz. tPA, fXa, thrombin, plasmin, and trypsin. CoMFA steric and electrostatic contour plots developed, signify the effect of various substituents on the activity, which can be employed for the rational design of novel selective serine protease inhibitors.

2. Results and discussion

3D-QSAR models for indole/benzoimidazole-5-carboxamidines as trypsin-like serine protease inhibitors (tPA, fXa, thrombin, plasmin, and trypsin) were derived from CoMFA methodology using in vitro inhibitory activity pK_i (μM) as a dependent variable. The development of predictive CoMFA models is essentially alignment sensitive that defines the putative pharmacophore for the series of ligands under investigation. The CoMFA models generated for indole/benzoimidazole-5-carboxamidines as uPA inhibitors using atom-based RMS fit method yielded better predictive models than shape-based RMS fitting, flexible multifit and RMSD based database fitting techniques.[27] The low-energy conformer obtained from systematic search routine was used for molecular superimposition of ligands on the template structure (compound **27**) by atom-based RMS fitting (Fig. 1). The structures and biological activity of training and test set molecules employed in generation and validation of 3D-QSAR models are given in Tables 1 and 2, respectively. σ minimum of 1.0 kcal/mol was used as threshold column filtering value in PLS analysis. The contribution of individual CoMFA fields in activity was studied by performing PLS analysis for only steric and electrostatic field. The in vitro enzyme inhibition data (K_i μM) used in the study could have contributions not only from steric and electrostatic fields but also from



27

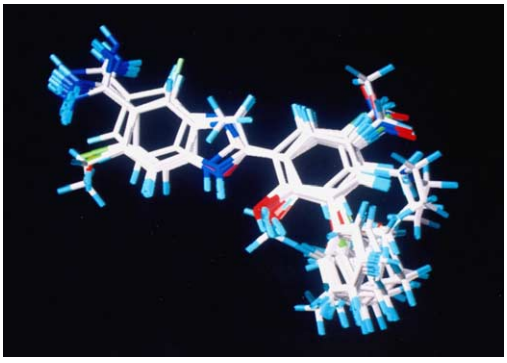


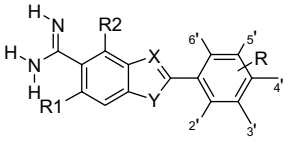
Figure 1. Atoms used in the superimposition of molecules on the template structure (compound **27**) by the RMS fit method. The aligned molecules are shown in capped sticks.

other physicochemical properties particularly hydrophobicity of the molecules accounting for their transport phenomenon and pharmacokinetic profile. To study the effect of hydrophobic/lipophilic properties of molecules on the activity, we have included ClogP as an additional descriptor in the CoMFA analysis. The CoMFA models obtained by inclusion of ClogP were statistically significant with minimum contribution of lipophilic parameters compared to steric and electrostatic fields. The results of PLS analyses are reported in Table 3.

2.1. tPA CoMFA model

The CoMFA model for tPA was obtained from atom-based RMS alignment using 28 training set molecules

Table 1. The structures and calculated ClogP of the training and test set molecules



Compound	X	Y	R					R1	R2	ClogP
			2′	3′	4′	5′	6′			
1	N	NH	OH	H	H	H	H	H	H	1.5888
2	N	NH	OH	H	H	F	H	H	H	1.9067
3	N	NH	OH	H	H	Br	H	H	H	2.6267
4	N	NH	OH	H	H	Me	H	H	H	2.0878
5	N	NH	OH	H	H	OMe	H	H	H	1.6621
6	N	NH	OH	NO_2	H	H	H	H	H	1.839
7	N	NH	OH	F	H	H	H	H	H	1.7067
8	N	NH	OH	Br	H	H	H	H	H	2.3467
9	N	NH	OH	OMe	H	H	H	H	H	1.4121
10	N	NH	OH	Ph	H	H	H	H	H	2.9768
11	N	NH	OH	H	Me	H	H	H	H	2.0878
12	N	NH	OH	H	H	H	OH	H	H	1.1126
13	N	NH	OH	-naphthyl-		H	H	H	H	2.7628
14	N	NH	OH	Br	H	Me	H	H	H	2.8457
15	CH	NH	OH	H	H	H	H	H	H	1.626
16	CH	NH	OH	Br	H	Me	H	H	H	3.0095
17	CH	NH	OH	Br	H	H	H	H	H	2.5105
18	CH	NH	OH	Br	H	NO_2	H	H	H	2.6833
19	CH	NH	OH	Ph	H	Me	H	H	H	3.513
20	CH	NH	OH	Ph	H	NO_2	H	H	H	3.3828
21	CH	NH	H	Ph	H	H	H	H	H	4.681
22	CH	NH	OMe	Ph	H	H	H	H	H	3.4895
23	NMe	N	OH	Ph	H	H	H	H	H	2.9568
24	CH	O	OH	Ph	H	H	H	H	H	3.524
25	CH	NH	OH	Ph	H	H	H	F	H	3.3322
26	CH	NH	OH	Ph	H	H	H	OMe	H	3.0390
27	CH	NH	OH	Ph	H	H	H	H	H	3.014
28	CH	NH	OH	Ph	H	H	H	Cl	H	3.9022
29	CH	NH	OH	Ph	H	H	H	OH	H	2.411
30	CH	NH	OH	Ph	H	H	H	H	Cl	3.9022
31	N	NH	OH	H	H	NO_2	H	H	H	1.8390
32	N	NH	OH	H	NEt_2	H	H	H	H	2.9216
33	N	NH	OH	Ph	H	Cl	H	H	H	3.8647
34	CH	NH	OH	Br	H	Br	H	H	H	3.4790
35	CH	NH	OH	Ph	H	Br	H	H	H	4.1785
36	CH	NH	OH	Ph	H	Cl	H	H	H	4.0285
37	CH	NH	OH	Ph	H	H	H	Me	H	3.513
38	N	NH	H	H	H	H	H	H	H	2.565
39	N	NH	OH	Me	H	H	H	H	H	2.0378

Table 2. Biological activities of the training and test set molecules ($\log 1/K_i$ μM)

Compound	tPA[a]		fXa[b]		Thrombin[c]		Plasmin[d]		Trypsin[e]	
	Actual	Calculated	Actual	Calculated	Actual	Calculated	Actual	Calculated	Actual	Calculated
1	−1.875	−1.632	−1.146	−1.008	−1.740	−1.917	−1.662	−1.486	−1.361	−1.265
2	−1.668	−1.653	−0.832	−1.231	−1.518	−2.070	−1.322	−1.588	−1.000	−1.402
3	−1.612	−1.767	−0.755	−0.768	−1.568	−1.725	−1.462	−1.421	−1.146	−1.223
4	−2.053	−1.880	−1.204	−0.525	−1.812	−1.452	−1.778	−1.230	−0.954	−1.004
5	−1.799	−1.779	−0.949	−0.758	−1.612	−1.647	−1.397	−1.414	−0.812	−1.207
6	−1.146	−1.171	−0.113	0.557	−0.041	0.358	−1.176	−0.699	−0.929	−0.431
7	−1.000	−1.113	0.050	−0.535	−0.845	−1.257	−0.740	−1.274	−0.544	−0.964
8	−0.531	−0.349	0.522	0.318	−0.361	−0.458	−0.230	−0.572	−0.146	−0.459
9	−1.838	−1.860	−1.204	−0.348	−1.903	−1.485	−1.278	−1.132	−1.477	−1.101
10	0	−0.025	−0.322	−0.527	−1.176	−1.393	−0.698	−0.769	−0.812	−0.851
11	−1.681	−1.794	−0.973	−0.944	−1.579	−1.787	−1.544	−1.435	−1.255	−1.190
12	−1.397	−1.424	−0.698	−0.873	−1.230	−1.772	−0.875	−1.407	−0.778	−1.239
13	−1.491	−1.513	−0.995	0.442	−1.903	−0.373	−1.518	−0.559	−1.113	−0.493
14	−0.380	−0.558	0.244	0.783	−0.431	0.005	−0.505	−0.327	−0.518	−0.207
15	−0.973	−1.001	−0.431	−0.564	−1.113	−1.392	−1.079	−0.987	−0.903	−0.810
16	0.337	0.483	1.283	1.333	0.744	0.805	0.221	0.217	0.494	0.364
17	0.420	0.390	1.366	0.773	0.920	0.183	0.602	−0.065	0.537	0.049
18	−0.113	−0.184	1.107	1.597	0.619	1.132	−0.322	0.078	−0.633	−0.144
19	0.958	1.050	0.677	0.880	−0.041	−0.032	0.602	0.824	0.886	0.992
20	1.721	1.799	1.853	1.312	0.958	0.692	0.721	0.684	1.124	0.678
21	−2.139	−2.105	−1.633	−1.369	−1.505	−1.373	−1.414	−1.353	−1.230	−0.908
22	−1.278	−1.295	−1.278	−1.142	−1.342	−1.092	−1.414	−1.268	−0.079	−0.221
23	−2.176	−2.515	−1.477	−1.925	−1.845	−1.683	−1.079	−1.697	−0.812	−1.033
24	−0.447	−1.603	−0.832	−1.053	−0.845	−0.931	−0.740	−1.176	−0.204	−0.332
25	−0.556	−0.802	−0.113	−0.338	−0.875	−1.039	0.455	−0.046	−0.041	−0.025
26	−1.255	−1.109	−1.414	−0.806	>−1.875	−1.086	−0.591	−0.087	−1.146	0.222
27	1.455	1.270	1.107	0.165	0.494	−0.694	1.000	0.356	0.886	0.353
28	−0.944	−1.013	−1.278	0.213	−1.778	−1.159	0.958	0.801	0.638	0.801
29	−1.662	−1.528	>−1.875	−0.382	>−1.875	−1.035	−1.278	−0.092	−1.698	−0.054
30	−0.518	0.587	0.716	−0.222	−1.812	−1.168	−0.477	−0.048	−0.672	−0.161
31	−1.748	−1.903	−0.892	−0.861	−1.531	−1.629	−1.612	−1.671	−1.477	−1.656
32	−2.000	−2.769	−1.397	−0.968	−1.954	−1.858	−1.903	−1.433	−1.204	−1.281
33	0.552	−0.043	0.026	−0.257	−1.079	−1.148	−0.612	−0.723	−0.518	−0.841
34	0.161	0.284	1.050	1.177	0.494	0.586	0.045	0.135	0	0.184
35	1.026	1.388	0.677	0.532	−0.146	−0.185	0.455	0.534	0.154	0.560
36	0.795	1.411	0.468	0.581	0.527	−0.101	0.537	0.521	0.886	0.513
37	−1.591	−0.680	>−1.875	0.156	>−1.875	−1.081	0.677	0.662	−0.204	0.715
38	−1.838	−0.186	−1.000	0.220	−1.681	−0.659	−1.778	−0.557	−1.176	−0.490
39	−2.580	−1.410	−2.096	−0.882	−2.531	−1.704	−2.544	−1.474	−1.740	−1.184

[a] Training set: 1–28; Test set: 29–36; Outliers: 37–39.
[b] Training set: 1, 2, 6–8, 10, 11, 15–24, 26, 27, 30, 31, 33, 36, 39; Test set: 4, 5, 12, 14, 25, 32, 35; Outliers: 3, 9, 13, 28, 29, 37, 38.
[c] Training set: 1–3, 6–11, 13, 15–24, 27, 28, 30, 31, 33, 36, 38, 39; Test set: 4, 5, 12, 14, 25, 32, 34; Outliers: 26, 29, 37.
[d] Training set: 1–3, 6–11, 15–24, 26–28, 30, 31, 33, 34, 36, 39; Test set: 4, 5, 12, 14, 25, 32, 35, 37; Outliers: 13, 29, 38.
[e] Training set: 1–3, 6–11, 15–24, 26, 28–31, 33, 35–37, 39; Test set: 4, 5, 12, 14, 25, 32, 34; Outliers: 13, 27, 38.

and validated by predicting activity for the eight test set molecules.

The model obtained with only the steric field showed cross-validated r^2 (r^2_{cv}) 0.692 with 10 components, non-cross-validated r^2 (r^2_{ncv}) 0.984, *F*-value 106.357, bootstrapped r^2 (r^2_{bs}) 0.992, while the model obtained with only the electrostatic field showed r^2_{cv} 0.199 with one component, r^2_{ncv} 0.403, *F*-value 17.577, r^2_{bs} 0.487.

The statistically significant model obtained with both steric and electrostatic fields showed r^2_{cv} 0.584 with seven components, r^2_{ncv} 0.987, *F*-value 224.836, r^2_{bs} 0.955, predictive r^2 (r^2_{pred}) 0.894, the steric and electrostatic contributions were 57.2% and 42.8%, respectively. The graph of actual versus fitted/predicted activities for the training/test set molecules is depicted in Figure 2. Figure 3 describes the CoMFA steric and electrostatic contour map for tPA inhibition.

The CoMFA steric map (Fig. 3) encompasses green contours (80% contribution) embedded in the 3′-phenyl ring and in the vicinity of 3-H of compound **27** reveals that increase in activity may be anticipated due to increased steric bulk, while the yellow contours (20% contribution) observed near 3′-phenyl ring, 6-H and 2′-OH of compound **27** suggests that steric substitutions at 3′-phenyl ring, C-6 (compound **26**), 2′-positions of compound **27** does not favor tPA inhibition.

The CoMFA electrostatic map (Fig. 3) displays red contours (80% contribution) near 4′,5′-protons, at the lower surface of 1-NH and surrounding the 3′-phenyl ring where the partial negative charge is associated with in-

Table 3. Summary of CoMFA results[a]

	tPA		fXa		Thrombin		Plasmin		Trypsin	
	SE	ClogP	SE	ClogP	SE	ClogP	SE	ClogP	SE	ClogP
r^2_{cv} [b]	0.584	0.530	0.513	0.509	0.504	0.488	0.629	0.576	0.583	0.443
Components	7	9	2	4	3	4	2	3	2	4
SEP[c]	0.809	0.899	0.798	0.840	0.757	0.787	0.583	0.637	0.558	0.675
r^2_{ncv} [d]	0.987	0.992	0.811	0.857	0.845	0.902	0.822	0.831	0.838	0.871
SEE[e]	0.140	0.115	0.497	0.425	0.423	0.344	0.404	0.402	0.348	0.325
F-value	224.836	258.363	47.252	34.843	39.962	48.416	57.788	39.242	59.497	35.447
$Pr^2_{=0}$	0	0	0	0	0	0	0	0	0	0
Contributions (%)										
Steric	57.2	51.7	56.1	50.0	54.5	56.4	61.8	57.9	58.2	56.4
Electrostatic	42.8	47.0	43.9	44.8	45.5	42.7	38.2	39.6	41.8	42.7
ClogP	—	1.3	—	5.2	—	0.9	—	2.4	—	0.8
r^2_{pred} [f]	0.894	0.737	0.753	0.851	0.860	0.724	0.857	0.854	0.621	0.566
r^2_{bs} [g]	0.955	0.998	0.819	0.924	0.896	0.922	0.842	0.858	0.833	0.930
Standard deviation[g]	0.003	0.002	0.054	0.029	0.036	0.030	0.053	0.044	0.055	0.031

[a] Alignment by atom-based RMS fit using both the steric and electrostatic fields.
[b] Cross-validated r^2 by leave-one-out method.
[c] Standard error of prediction.
[d] Non-cross-validated r^2.
[e] Standard error of estimate.
[f] Predictive r^2.
[g] From 100 bootstrapping runs.

creased activity (compound **20**), while blue contours (20% contribution) observed at the upper surface of

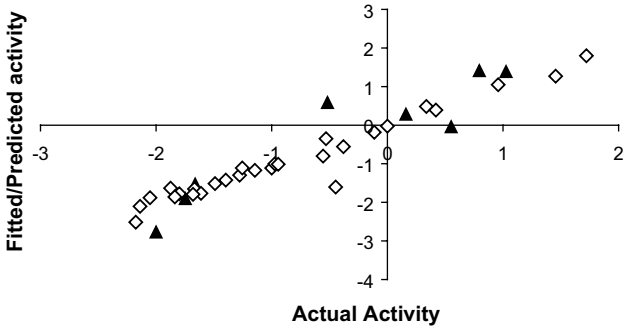


Figure 2. A graph of actual versus fitted/predicted activities (log $1/K_i$ μM) of the training/test set molecules as tPA inhibitors (◊ training set; ▲ test set).

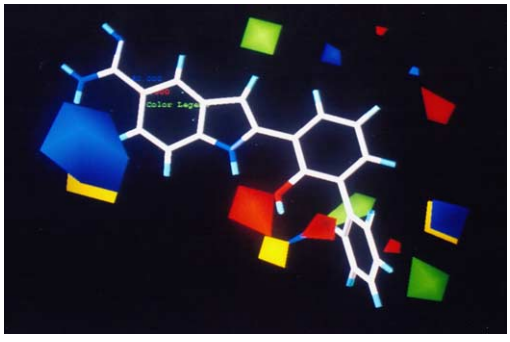


Figure 3. The CoMFA steric and electrostatic STDDEV*COEFF contour plots for tPA inhibition (green polyhedron: sterically favored, yellow polyhedron: sterically disfavored, red polyhedron: negatively charge favored, blue polyhedron: positively charge favored). The template structure (compound **27**) is shown in capped sticks.

2-phenyl ring and in the vicinity of 6-H of compound **27** indicates areas within the lattice where electropositive properties of the molecule describe an increase in activity.

2.2. fXa CoMFA model

Using 25 training set molecules, the CoMFA model for fXa was generated from atom-based RMS alignment and validated from an external test set comprising of seven molecules.

The model obtained from only the steric field showed r^2_{cv} 0.492 with three components, r^2_{ncv} 0.775, *F*-value 24.047, r^2_{bs} 0.855 whereas the model from the electrostatic field showed r^2_{cv} 0.622 with five components, r^2_{ncv} 0.956, *F*-value 82.442, r^2_{bs} 0.966. The CoMFA model with both steric and electrostatic fields showed r^2_{cv} 0.513 from first two components, r^2_{ncv} 0.811, *F*-value 47.252, r^2_{bs} 0.819, r^2_{pred} 0.753 with 56.1% steric and 43.9% electrostatic field contributions. The graph of actual versus fitted/predicted activity for the training/test set molecules is depicted in Figure 4. The steric and electrostatic contour maps for the same are shown in Figure 5.

The CoMFA steric map (Fig. 5) displays a favorable green polyhedron embedded in the lower part of 3′-phenyl ring and in the vicinity of 3-H, whereas the sterically unfavorable large yellow polyhedron was observed surrounding the lower region of 3′-phenyl ring and small contour flying near 2′-OH of compound **27**.

The CoMFA electrostatic map (Fig. 5) showed a positively charge favorable blue contour near 6-H and a large cavity engulfing 2-phenyl ring and 2′-OH position, the negatively charge favorable red contours were placed

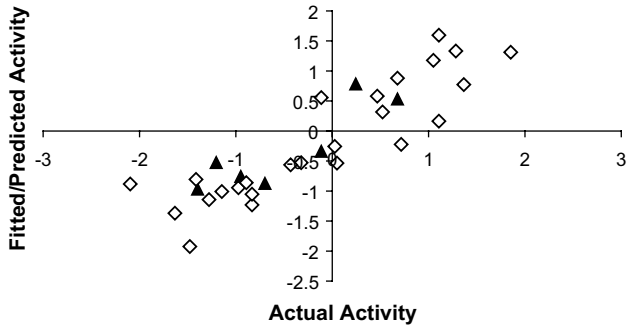


Figure 4. A graph of actual versus fitted/predicted activities (log $1/K_i$ μM) of the training/test set molecules as fXA inhibitors (◊ training set; ▲ test set).

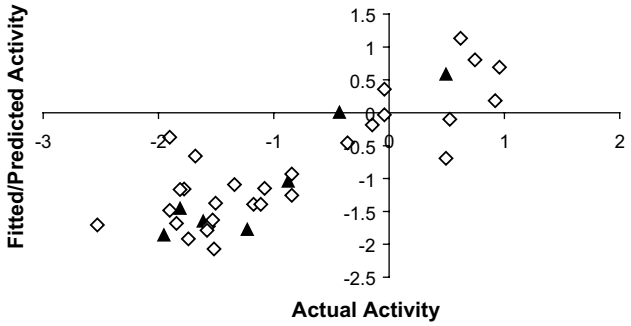


Figure 6. A graph of actual versus fitted/predicted activities (log $1/K_i$ μM) of the training/test set molecules as thrombin inhibitors (◊ training set; ▲ test set).

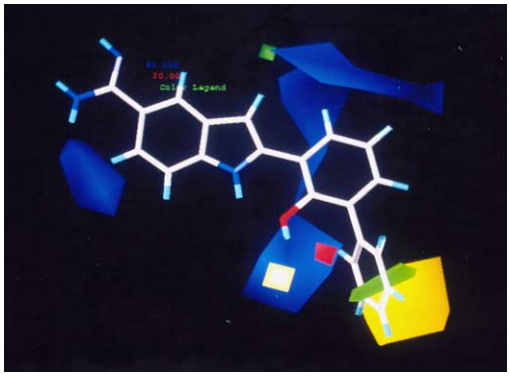


Figure 5. The CoMFA steric and electrostatic STDDEV*COEFF contour plots for fXa inhibition (green polyhedron: sterically favored, yellow polyhedron: sterically disfavored, red polyhedron: negatively charge favored, blue polyhedron: positively charge favored). The template structure (compound **27**) is shown in capped sticks.

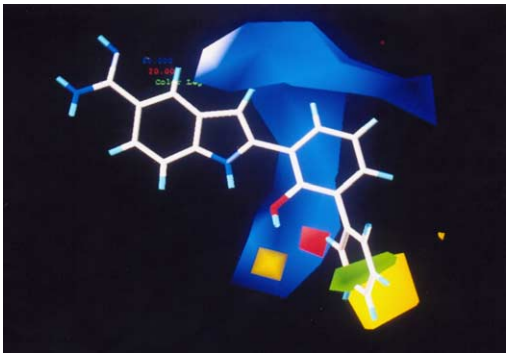


Figure 7. The CoMFA steric and electrostatic STDDEV*COEFF contour plots for thrombin inhibition (green polyhedron: sterically favored, yellow polyhedron: sterically disfavored, red polyhedron: negatively charge favored, blue polyhedron: positively charge favored). The template structure (compound **27**) is shown in capped sticks.

in the region between 2′-OH and 3′-phenyl ring of compound **27**.

2.3. Thrombin CoMFA model

The 3D-QSAR CoMFA model for thrombin was generated using 29 training set molecules and validated with seven test set molecules.

The CoMFA model obtained with only the steric field showed r^2_{cv} 0.520 from nine components, r^2_{ncv} 0.969, *F*-value 55.811, r^2_{bs} 0.985 while the model with only the electrostatic field showed r^2_{cv} 0.425 from five components, r^2_{ncv} 0.992, *F*-value 51.296, r^2_{bs} 0.958. The CoMFA model generated from both steric and electrostatic fields exhibited r^2_{cv} 0.504 with three components, r^2_{ncv} 0.845, *F*-value 39.962, r^2_{bs} 0.896, r^2_{pred} 0.860 with 54.5% steric and 45.5% electrostatic field contributions. The graph of actual versus fitted/predicted activities for the training/test set molecules is depicted in Figure 6. The CoMFA steric and electrostatic contour maps developed for the above validated model is shown in Figure 7.

The CoMFA steric contour map for thrombin (Fig. 7) is placed almost similar to tPA with sterically favorable green polyhedron observed in the vicinity of 3-H and embedded in 3′-phenyl ring, while the unfavorable yellow contours were in the lower region of 3′-phenyl ring and near 2′-OH of compound **27**.

The CoMFA electrostatic contour map (Fig. 7) showed a large blue cavity extending from the upper surface of 2-phenyl ring and negatively charge favored red contours between 2′-OH and 3′-phenyl ring almost similar to fXa.

2.4. Plasmin CoMFA model

The CoMFA model for plasmin was generated using 28 training set molecules and validated by predicting the activities of eight test set molecules.

The model obtained with only the steric field showed r^2_{cv} 0.614 with three components, r^2_{ncv} 0.822, *F*-value 36.996, r^2_{bs} 0.883, while the model from only electrostatic field showed r^2_{cv} 0.646 with 10 components, r^2_{ncv} 0.986, *F*-value 118.912, r^2_{bs} 0.995. The CoMFA model for both steric and electrostatic fields yielded r^2_{cv} 0.629 with first two components, r^2_{ncv} 0.822, *F*-value 57.788, r^2_{bs} 0.842, r^2_{pred} 0.857 with 61.8% steric and 38.2% electrostatic field contributions. Figure 8 describes the graph of actual versus fitted/predicted activities for the training/test set molecules. The steric and electrostatic contour maps gener-

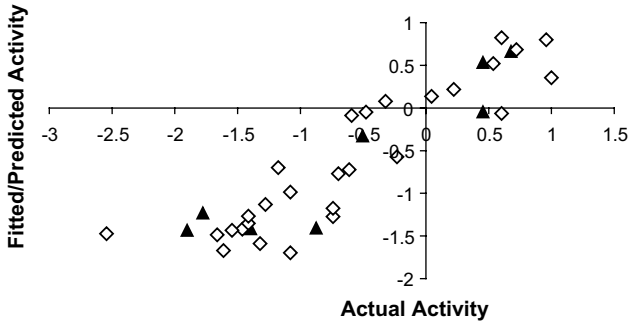


Figure 8. A graph of actual versus fitted/predicted activities ($\log 1/K_i$ μM) of the training/test set molecules as plasmin inhibitors (◊ training set; ▲ test set).

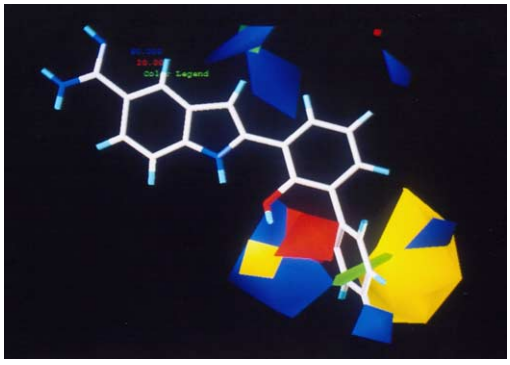


Figure 9. The CoMFA steric and electrostatic STDDEV*COEFF contour plots for plasmin inhibition (green polyhedron: sterically favored, yellow polyhedron: sterically disfavored, red polyhedron: negatively charge favored, blue polyhedron: positively charge favored). The template structure (compound **27**) is shown in capped sticks.

ated for the CoMFA model is shown in Figure 9 and were placed almost similar to fXa inhibition.

The CoMFA steric map (Fig. 9) for plasmin inhibitors encompasses of sterically favorable green contours embedded in the lower region of 3′-phenyl ring and in the vicinity of 3-H, whereas the sterically unfavorable yellow contours surround the lower region of 3′-phenyl ring and near 2′-OH of compound **27**.

The CoMFA electrostatic map (Fig. 9) displays blue contours at the upper surface of 2-phenyl ring, the space between 1-NH and 2′-OH of compound **27** where increase in activity may be due to presence of positively charge substituents. The negatively charge favorable red contours were found in the space between 2′-OH and 3′-phenyl ring of compound **27.**

2.5. Trypsin CoMFA model

The CoMFA model for trypsin inhibition was generated using 29 training set molecules and validated by predicting the activity of seven test set molecules.

The CoMFA model obtained with only the steric field yielded r^2_{cv} 0.583 with three components, r^2_{ncv} 0.813, *F*-value 31.884, r^2_{bs} 0.866, while the model obtained from only the electrostatic field showed r^2_{cv} 0.522 at first component, r^2_{ncv} 0.664, *F*-value 47.499, r^2_{bs} 0.694.

The statistically significant CoMFA model for trypsin inhibition obtained from both the steric and electrostatic fields showed r^2_{cv} 0.583 with two components, r^2_{ncv} 0.838, *F*-value 59.497, r^2_{bs} 0.833, r^2_{pred} 0.621, the steric and electrostatic field contributions were 58.2% and 41.8%, respectively. The graph of actual versus fitted/predicted activities for the training/test set molecules is depicted in Figure 10. The CoMFA steric and electrostatic contour maps developed for trypsin inhibition is shown in Figure 11.

The CoMFA steric map (Fig. 11) displayed the sterically favorable green polyhedron embedded in 3′-phenyl ring and in the vicinity of 3-H of compound **27** where optimum steric substituents may enhance the activity and the sterically disfavorable yellow polyhedron was placed similar to fXa, plasmin, and thrombin inhibition.

The CoMFA electrostatic map (Fig. 11) showed positively charge favored large blue polyhedron at the upper surface of 2-phenyl ring, small blue contours near the lower surface of 3′-phenyl ring and negative charge favored red contours in the space between 2′-OH and 3′-phenyl ring of compound **27**.

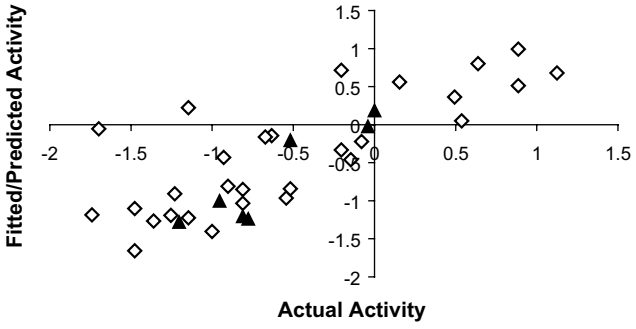


Figure 10. A graph of actual versus fitted/predicted activities ($\log 1/K_i$ μM) of the training/test set molecules as trypsin inhibitors (◊ training set; ▲ test set).

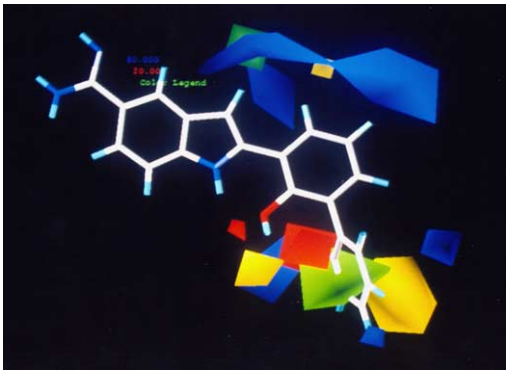


Figure 11. The CoMFA steric and electrostatic STDDEV*COEFF contour plots for trypsin inhibition (green polyhedron: sterically favored, yellow polyhedron: sterically disfavored, red polyhedron: negatively charge favored, blue polyhedron: positively charge favored). The template structure (compound **27**) is shown in capped sticks.

2.6. uPA CoMFA model

The detailed CoMFA results for uPA inhibition has been reported in our earlier publication.[27] The statistically significant CoMFA model (r^2_{cv} 0.611, r^2_{cnv} 0.778, *F*-value 43.825, r^2_{bs} 0.842, r^2_{pred} 0.616 with two components) was established from 28 molecules and validated by predicting the activities of nine test set molecules.

The CoMFA contour map displayed sterically favorable green contours embedded in the 3′-phenyl ring and in the vicinity of 3-H, whereas the disfavored yellow contours were observed bordering the 3′-phenyl ring of compound **27**. The electrostatic contour map displayed negative charge favored red contours in the vicinity of 6-H and in the region between 2′-OH and 3′-phenyl ring, while the positively charge favored large blue contours were displayed in the upper portion of 2-phenyl ring of the compound **27**.

In the present study, we have developed statistically significant predictive CoMFA models using a series of indole/benzoimidazole-5-carboxamidines for various serine protease inhibitors employing steric and electrostatic fields. The generated CoMFA models exhibited higher contributions of steric parameters toward the activity of these molecules.

An inspection of the fitted/predicted activities for the data set revealed poor prediction for some of the indoles (compounds **13**, **38**, and **39**) and benzoimidazoles (compounds **17**, **26**, **29**, **30**, and **37**) considered as the outliers. The outliers for the indole series: compound **13** was the only 2-naphthyl derivative, compound **38** does not possess the hydroxyl group at 2′-position and compound **39** was the only compound with steric 3′-methyl group. In case of benzoimidazoles, compound **17** was the only 3′-Br derivative in the sterically favored region, compounds **26** and **37** had methoxy and methyl groups at sixth position, respectively, where electropositive substituents favor activity. The other outliers in this series compound **29** was the only compound with lipophilic hydroxyl group at sixth position and compound **30** as the only 4-Cl derivative. The structural uniqueness and composition of training set molecules might be the reason for poor prediction. The compounds for which activity is not described in specific numeric terms and the poorly predicted molecules from the non-cross-validated PLS analysis have been termed as outliers in the present study.

All the steric and electrostatic contour maps generated from the validated CoMFA models for the serine protease inhibitors were used to analyze the selectivity trends of indole/benzoimidazole-5-carboxamidines. The steric contour maps displayed sterically disfavored yellow contours in the lower region of 3′-phenyl ring and in the vicinity of 2′-OH of compound **27** except for tPA. The sterically favorable green contours were observed near 3-H and embedded in 3′-phenyl ring of compound **27** except for tPA, where the green contours appeared surrounding 3′-phenyl ring. Comparison of the green polyhedron in the lower region of 3′-phenyl ring showed some degree of differences for all the serine protease inhibitors suggesting that optimum steric substitution at 3′-C of 2-phenyl ring may play a crucial role in achieving improved activity.

Electrostatic contour maps displayed negatively charge favorable red polyhedron in the regions between 2′-OH and 3′-phenyl ring except for tPA, whereas the positively charge favorable blue contours were observed at the upper surface of 2-phenyl ring. These regions include the partial positive charges associated with hydrogen atoms bound to carbon, which can be correlated with lipophilic interactions.

Additionally, tPA and fXa displayed blue contour in the vicinity of 6-H where optimum low electron density substituents may be crucial in selectivity of serine protease inhibitors. The indole derivatives (compounds **8** and **10**) with similar substituents showed less activity than benzoimidazole derivatives (compounds **17** and **27**). The replacement of 2′-OH group with H (compounds **21** and **38**), OMe (compound **22**) and 1-NH with N (compound **23**) and O (compound **24**) resulted in reduced activity, demonstrating the significance of 2′-OH and 1-NH groups in binding the receptor. Overall in the present study, the CoMFA models developed for the series of indole/benzoimidazole-5-carboxamidines as serine protease inhibitors highlight important structural requirements, which can be used in designing selective serine protease inhibitors.

3. Conclusion

Receptor independent 3D-QSAR/CoMFA models using a series of indole/benzoimidazole-5-carboxamidines for various trypsin-like serine protease inhibitors viz. tPA, fXa, thrombin, plasmin, and trypsin have been established. All the molecular modeling and 3D-QSAR studies were performed with the standard protocol using SYBYL 6.7 software.

As can be seen from Table 3, reasonably correlative and predictive CoMFA models were generated from the combination of steric and electrostatic fields for various serine protease inhibitors than the CoMFA models generated using ClogP as an additional descriptor. This shows that variations in observed biological activity can be explained from steric and electrostatic field interactions of the compounds. PLS analysis with the inclusion of ClogP resulted in a slightly decreased cross-validated r^2 value. The ClogP contribution in these models varied from 0.8–5.2%, hence lipophilicity may not be a determining factor for the activity of indole/benzoimidazole-5-carboxamidines as serine protease inhibitors. The statistically significant and validated 3D-QSAR/CoMFA models obtained from steric and electrostatic fields were thus employed to generate 3D contour plots for various serine protease inhibitors.

The comparison of steric and electrostatic contour maps showed high similarity in catalytic/hydrophobic site

(near C-3′ substitution). In all the cases, the sterically favorable green contours were in the vicinity of C-3 substitution and embedded in 3′-phenyl ring, whereas in tPA they were observed adjoining the 3′-phenyl ring. The sterically disfavored yellow contours were observed in the vicinity of 3′-phenyl ring of compound **27**.

The electrostatic contour plots displayed red polyhedron amid the 2′-OH and 3′-phenyl ring suggesting that increase in activity may be anticipated by moderate electronegative substitutions at 3′ position of 2-phenyl ring (compounds **7**, **8**, and **17**), while moderate electropositive substituents at C-3, 2′, 6′-positions may enhance the tPA, fXa, plasmin and thrombin activity as shown by the presence of positively charge favored blue contours.

In accordance with the findings of Verner et al.[28] and Mackman et al.[29] we have also observed that the nature of the substituent at C-6 position is the crucial factor in determining the selectivity trends of various serine protease inhibitors. Previously, we have reported that increase in uPA activity/selectivity of indole/benzoimidazole-5-carboxamidines can be anticipated by moderate electronegative substitution at C-6 position.[27] In the present study, we observed that optimum low electron density substitutent at C-6 may enhance the tPA and fXa activity as evident from the presence of the blue contour near 6-H of compound **27**. Replacement of C-6 proton with a halogen (chloro) markedly decreased the tPA, fXa and thrombin inhibition, whereas no significant change was observed in case of plasmin and trypsin (compound **28**). 3D-QSAR/CoMFA models developed for indole/benzoimidazole-5-carboxamidines as trypsin-like serine protease inhibitors may provide further insight into the structural determinants, which can be exploited in the rational design of selective serine protease inhibitors for cancer as well as anticoagulant or antithrombotic chemotherapeutics.

4. Methods

4.1. Biological data

The data set consists of a series of indole/benzoimidazole-5-caboxamidines published in the literature as trypsin-like serine protease inhibitors viz. uPA, tPA, fXa, thrombin, plasmin, and trypsin.[28,29] The negative logarithm of measured K_i (μM) against serine proteases enzyme as pK_i (μM, $pK_i = \log 1/K_i$) was used in 3D-QSAR studies. Table 1 describes the structures and calculated ClogP values of the data set and Table 2 biological activity for the training/test set molecules. The robustness and predictive ability of models were evaluated by analyzing test set molecules with a wide range of biological activity.

4.2. Molecular modeling

All molecular modeling and 3D-QSAR studies were performed using SYBYL 6.7[30] with standard TRIPOS force field[31] running on Silicon Graphics Indy o2 workstation (Tripos Inc.). The molecules were constructed using standard geometrics and bond lengths with SYBYL. Initial optimization was carried out using standard TRIPOS force field employing Gasteiger Marsili charges, the constraints were removed and repeated minimization was performed using steepest descent and conjugated gradient method till the root mean square (rms) deviation of 0.001 kcal/mol was achieved. Systematic search routine was used in the conformational analysis and all rotatable bonds were searched in 10° increments from 0° to 359°. Conformational energies were computed with electrostatic term and the low-energy conformers were selected for superimpositions. Further optimization was performed using MOPAC with AM1 Hamiltonian[32] and MOPAC charges derived were used in subsequent analysis. Superimposition of the molecules carried out by atom-based RMS fitting on the template structure (compound **27)** was used in CoMFA probe interaction energy calculations. The superimposed molecules and atoms used in the alignment are depicted in Figure 1.

4.3. CoMFA 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 intersection with maximal steric interactions. The CoMFA fields generated automatically were scaled by CoMFA-STD method in SYBYL.

4.4. Partial least square (PLS) analysis

The CoMFA descriptors served as independent variables and pK_i (μM) values as dependent variables in deducing 3D-QSAR models by PLS regression analysis method.[33] The predictive values of models were evaluated by leave-one-out (LOO) cross validation method. The optimum number of components (N_c) used to derive the non-cross-validated model was defined as the number of components leading to the highest cross-validated r^2 and the lowest standard error of prediction (SEP). PLS analysis using 100 bootstrap groups with optimum number of components was performed in order to obtain the statistical confidence limit in the analyses.

4.5. Predictive r^2 values

The predictive ability of each analysis was determined from the test set that was 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 is defined as

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

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.

4.6. Calculation of ClogP

ClogP was calculated using ChemDraw Ultra 6.0 software integrated with CambridgeSoft Software Development Kit.[34]

Acknowledgements

We thank Dr. F. V. Manvi, Principal, Prof. A. D. Taranalli, Vice-Principal, KLES's College of Pharmacy, Belgaum, India, for providing necessary facilities. B.A.B. is thankful to Human Resource Development Group, Council for Scientific and Industrial Research, New Delhi for the award of Senior Research Fellowship [8/470(1)/2001-EMR-I] and Christopher G. Barber, Pfizer Global Research and Development, Sandwich, Kent CT 13 9NJ, UK for his valuable suggestions. We 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

1. Dano, K. *Adv. Cancer Res.* **1985**, *44*, 139.
2. Fazioli, F.; Blasi, F. *Trends Pharmacol. Sci.* **1994**, *15*, 25.
3. Evans, D. M.; Sloan-Stakleff, K. D. *Drug News Perspect.* **1997**, *10*, 85.
4. Collen, D.; Lijnen, H. R. *Blood* **1991**, *78*, 3114.
5. Madison, E. L. *Fibrinolysis* **1994**, *8*, 221.
6. Collen, D.; Lijnen, H. R. *Thromb. Haemostasis* **1995**, *74*, 161.
7. Cho, S. H.; Tam, S. W.; Demissic-Sanders, S.; Filler, S. A.; Oh, C. K. *J. Immunol.* **2000**, *165*, 3154.
8. Reich, R.; Thompson, E. W.; Iwamoto, Y.; Martin, G. R.; Deason, J. R.; Fuller, G. C.; Miskin, R. *Cancer Res.* **1988**, *48*, 3307.
9. Gladson, C. L.; Pijuan-Thompson, V.; Olman, M. A.; Gillespie, G. Y.; Yacoub, I. Z. *Am. J. Pathol.* **1995**, *146*, 1150.
10. Hsu, D. W.; Efird, J. T.; Hedley-Whyte, E. T. *Am. J. Pathol.* **1995**, *147*, 114.
11. Jankum, J.; Keck, R. W.; Skrzypczak-Jankum, E.; Swiercz, R. *Cancer Res.* **1997**, *57*, 559.
12. Mohanam, S.; Sawaya, R.; McCutcheon, I.; Ali-Osman, F.; Boyd, D.; Rao, J. S. *Cancer Res.* **1993**, *53*, 4143.
13. Ossowski, L.; Reich, E. *Cell* **1983**, *35*, 611.
14. Duffy, M. J. *Clin. Cancer Res.* **1996**, *2*, 613.
15. Magill, C.; Katz, B. A.; Mackman, R. L. *Emerg. Ther. Targets* **1999**, *3*, 109.
16. Colman, R. W.; Marder, V. J.; Salzman, E. W.; Hirsh, J. Overview of Hemostasis. In *Hemostasis and Thrombosis: Basic Principles and Clinical Practice*; Colman, R. W., Marder, V. J., Salzman, E. W., Eds.; J.B. Lippincott: Philadelphia, 1994; p 3.
17. Elodi, S.; Vardi, K. *Thromb. Res.* **1979**, *15*, 617.
18. Hoffman, M.; Monroe, D. M. *Thromb. Haemost.* **2001**, *85*, 958.
19. Bode, W.; Mayr, I.; Baumann, U.; Huber, R.; Stone, S. R.; Hofsteenge, J. *EMBO J.* **1989**, *8*, 3467.
20. Padmanabhan, K.; Padmanabhan, K. P.; Tulinsky, A.; Park, C. H.; Bode, W.; Huber, R.; Blankenship, D. T.; Cardin, A. D.; Kisiel, W. *J. Mol. Biol.* **1993**, *232*, 947.
21. Lamba, D.; Bauer, M.; Huber, R.; Fischer, S.; Rudolph, R.; Kohnert, U.; Bode, W. *J. Mol. Biol.* **1996**, *258*, 117.
22. Spraggon, G.; Phillips, C.; Nowak, U. K.; Ponting, C. P.; Saunders, D.; Dobson, C. M.; Stuart, D. I.; Jones, E. Y. *Structure* **1995**, *3*, 681.
23. Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
24. Katz, B. A.; Mackman, R.; Luong, C.; Radika, K.; Martelli, A.; Sprengeler, P. A.; Wang, J.; Chan, H.; Wong, L. *Chem. Biol.* **2000**, *7*, 299.
25. Rai, R.; Sprengeler, P. A.; Elrod, K. C.; Young, W. B. *Curr. Med. Chem.* **2000**, *1*, 35.
26. Fevig, J. M.; Wesler, R. R. *Annu. Rep. Med. Chem.* **1999**, *34*, 81.
27. Bhongade, B. A.; Gadad, A. K. *Bioorg. Med. Chem.* **2004**, *12*, 2797.
28. 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.
29. 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.
30. Sybyl 6.7, Tripos Associates, 1699 South Hanley Road, St. Louis, MO 63144.
31. Clark, M.; Cramer, R. D., III; Van, O. N. *J. Comput. Chem.* **1989**, *10*, 982.
32. Stewart, J. J. P. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1.
33. Wold, S.; Albano, C.; Dunn, W.; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjosstrom, M. In *Chemometrics: Mathematics and Statistics in Chemistry*; Kowalski, B., Ed.; Reidel: Dordrecht, The Netherlands, 1987; p 17.
34. ChemDraw Ultra 6.0, CambridgeSoft Corporation, 100 CambridgePark. Cambridge, MA 02140, USA.