

Pharmacophore Modeling and Docking Based QSAR Studies of Aryl Amidino Isoxazoline Derivatives to Design Potential FXa Inhibitors

Suresh B Vepuri^{1,*}, S. Anbazhagan², Naresh P¹, D. Divya¹

¹Institute of Pharmacy, GITAM University, Vishakapatnam, AP, 530045, India

²Karuna College of Pharmacy, Thirumittacode, Palakad, Kerala, 679533, India

Abstract In order to elucidate the structural requirements for human factor Xa receptor antagonism, 72 antagonists belonging to isoxazolidine chemical class were selected from the literature and conducted molecular modeling studies. Best binding conformations were isolated by docking selected molecules into the receptor binding site. To further explore the structure-activity relationships within the considered chemical class, a pharmacophore model and QSAR analyses were developed. Pharmacophore models of these inhibitors were established by using the HipHop and HypoGen algorithms implemented in the Catalyst software package. The best quantitative pharmacophore model, which has the highest correlation coefficient (0.92), consisted of two hydrogen bond donor, a hydrophobic aromatic, and a ring aromatic feature. The model was further validated by test set and cross validation method. Molecular shape analysis (MSA) and Molecular field analysis (MFA) were used as the QSAR techniques. Two conformer-based alignment strategies were employed to construct reliable 3D-QSAR models. The docked conformer-based alignment strategy gave the best 3D-QSAR models. The best MFA and MSA models gave a cross-validated coefficient $q(2)$ of 0.641 and 0.816, non-cross-validated $r(2)$ values of 0.736 and 0.902, 20% out $r(2)$ values of 0.743 and 0.871, respectively. The information obtained from molecular modelling studies was very helpful to design some novel selective inhibitors of factor Xa with desired activity.

Keywords Catalyst, Factor Xa, Ligandfit, MSA, MFA, Pharmacophore and QSAR

1. Introduction

Over the past twenty to thirty years, scientists have used computer models of new chemical entities to help define activity profiles, geometries and reactivities[1]. Computational chemistry/molecular modeling is the science (or art) of representing molecular structures numerically and simulating their behavior with the equations of quantum and classical physics[2]. The computational approaches used to design the drugs are dependent upon the amount of information that is available about the ligand and receptor. Based on the information that is available, one can apply either structure-based or ligand-based molecular design methods[3].

To discover novel ligands for receptors of known structure, investigators often use docking computer programs to screen multi-compound databases for molecules that fit a binding site on the receptor. Ligandfit[4] is one of such a drug discovery software program. QSAR and pharmacophore

techniques are ligand-based approaches used to correlate the biological activity with chemical structure of ligands. In QSAR studies, quantitative correlation of molecular structures with the activity/property, and subsequently, the prediction of this property for novel compounds is done. Furthermore, these methods help to characterize those spatial features that are responsible for activity changes in a series of active molecules[5]. Molecular Field Analysis (MFA)[6] and Molecular Shape Analysis (MSA)[6] are two different modules of QSAR study. MFA quantifies the interaction energy between a probe molecule and a set of aligned target molecules in a QSAR. Interaction energies measured and analyzed for a set of 3D structures can be useful in establishing QSARs whereas, MSA extends QSAR operations for performing 3D QSAR studies. This technique generates quantitative measurements of molecular shape properties as part of QSAR analysis[6]. Statistical methods are an essential component of QSAR work. They help to build models, estimate a model's predictive abilities, and find relationships and correlations among variables and activities. Regression methods are used to build a model in the form of an equation that gives one or more dependent variables (usually activity) in terms of independent variables ("descriptors"). The model can then be used to predict

* Corresponding author:

vsb.gip@gitam.in (Suresh B Vepuri)

Published online at <http://journal.sapub.org/bioinformatics>

Copyright © 2012 Scientific & Academic Publishing. All Rights Reserved

activities for new molecules, perhaps prioritizing or screening a large group of molecules whose activities are not known

A *pharmacophore* is defined as an ensemble of universal chemical features that characterize a specific mode of action of a ligand in the active site of the macromolecule in 3D space[7]. Chemical features are e.g. hydrogen bonds, charge interactions, hydrophobic areas. CATALYST[8] is a pharmacophore generating modeling tool. It has been used successfully, in conjunction with traditional research techniques, to examine the structural properties of existing compounds, develop and quantify a hypothesis which relates these properties to observed activity and utilize these "rules" to predict properties and activities for new chemical entities[8]. Catalyst tools help to rationally design small molecules as drug candidates using 3D pharmacophore and shape-based models, and to suggest potentially active compounds suitable for synthesis and biological testing

The clinical usefulness of anticoagulants such as warfarin (Coumadin) and low molecular weight heparins (LMWHs) have generated a great interest in searching for new anticoagulants for the treatment and prevention of thromboembolic diseases. However, warfarin is an indirect inhibitor of blood coagulation and takes several days to reach effective anticoagulant levels[9]. LMWH requires a physiological cofactor, antithrombin III, to inactivate factor Xa (fXa). Therefore, the current drug discovery research focuses on designing small molecule inhibitors that directly act on coagulation factors such as thrombin and fXa.

The direct thrombin inhibitors have shown good antithrombotic efficacy in animal thrombosis models. However, it is not certain whether the antithrombotic effects of these inhibitors can be achieved without bleeding complications in humans[10]. The alternative target of new anticoagulants is inhibition of the earlier sites of the coagulation cascade such as fXa. Factor Xa (fXa), a trypsin-like serine protease, is a critical enzyme which in combination with fVa and Ca^{2+} on a phospholipid surface forms the prothrombinase complex that converts biologically inactive prothrombin to thrombin. The activation of thrombin by fXa is a highly amplified process. Because of its central position at the convergent point of the intrinsic and extrinsic pathways of coagulation it is believed that inhibition of fXa may reduce the production of thrombin by either the extrinsic or intrinsic pathways without interfering with a basal level of thrombin activity necessary for normal hemostasis. The present study involves estimation of structure–activity relationship of aryl amidino isoxazoline derivatives which were reported as fXa inhibitors. This study would help in the rational development of new, synthetic, potent fXa inhibitors since it gives an insight into structural requirements for this series of inhibitors with the help of different QSAR methods.

2. Materials and Methods

2.1. Docking

All docking studies were performed using Accelrys Ligandfit installed on a Silicon Graphic Octane desktop Workstation. Shape and volume of the active site has been identified by selecting the aminoacid residues around the non stranded residue of the protein at 5 Å distance. A set of aryl amidino isoxazoline derivatives from literature was used in this study[11-15] (Table-1 to Table-5). Ten Molecules with varied range of activities were selected from the study set molecules. Conformational search on each selected molecules was done by simplex search method. Generated conformers were docked in the active site and calculated for their interactions. Ten best conformations for each molecule were selected by the system based on dock score and ligand interaction energies.

2.2. 3D-QSAR

Factor Xa binding affinity data reported by Quan et al. was used for the QSAR study. The affinity data[$\text{Ki}(\text{nM})$] of aryl amidino isoxazoline derivatives (Table 1-5) for factor Xa was converted to the logarithmic scale[pKi] and then used for subsequent QSAR analyses as the response variable. All computational experiments were conducted with Cerius2 4.11[16] version QSAR environment from Accelrys (San Diego, USA) on a Silicon Graphics O2 workstation running under the IRIX 6.5 operating system. Molecular shape analysis (MSA) and Molecular field analysis (MFA) were used as the 3D-QSAR techniques.

The major steps of MSA were (1) generation of conformers and energy minimization; (2) hypothesizing an active conformer (global minimum of the most active compound); (3) selecting a candidate shape reference compound (based on the active conformation); (4) performing pairwise molecular superimposition using the maximum common subgroup (MCSG) method; (5) measuring molecular shape commonality using MSA descriptors; (6) determining other molecular features by calculating quantum mechanical, spatial, electronic and conformational parameters; (7) selection of conformers; (8) generation of QSAR equations by genetic function algorithm (GFA) or stepwise regression. A complete list of descriptors used for the study molecules were presented in Table-7. Multiple conformations of each molecule were generated using the Boltzmann jump conformational search method. The upper limit of the number of conformations per molecule was 150. Each conformer was subjected to an energy minimization procedure using the smart minimizer with the Drieding force field to generate the lowest energy conformation for each structure. Varying the conformation of shape reference compound during alignment, two separate studies were conducted. In one study (MSA1) the best conformer of the most active inhibitor 51 was used as shape reference for the alignment. Best conformer was selected based on dock score and interaction energy which were obtained in flexible docking study (Ligandfit). In another study (MSA2) lowest energy conformer obtained in conformational search was used. A rigid fit of atom pairings

was performed to superimpose each structure so that it overlays the shape-reference compound.

The major steps of MFA were (1) generating conformers and energy minimization; (2) matching atoms using maximum common substructure (MCS) search and aligning molecules using the default options; (3) setting MFA preferences (rectangular grid with 2.00 Å step size, charges by Gasteiger algorithm, H⁺, CH₃ and HO⁻ as probes); (4) creating the field; (5) analysis by the Genetic partial least squares (G/PLS) method. The MFA formalism calculates probe interaction energies on a rectangular grid around a bundle of active molecules. The surface was generated from a "Shape Field". The atomic coordinates of the contributing models were used to compute field values on each point of a 3D grid. Grid size was adjusted to default 2.00 Å. The MFA evaluates the energy between a probe (H⁺, CH₃ and HO⁻) and a molecular model at a series of points defined by a rectangular grid. Fields of molecules were represented using grids in MFA and each energy associated with an MFA grid point serves as input for the calculation of a QSAR. These energies were added to the study table to form new columns headed according to the probe type. Statistical analysis of

data was done using techniques like G/PLS, genetic function approximation (GFA) using QSAR+ environment of Cerius2 software. The GFA technique was used to generate a population of equations rather than one single equation for correlation between biological activity and physicochemical properties. The GFA was done as followed: (1) an initial population of equations is generated by random choice of descriptors; (2) pairs from the population of equations were chosen at random and "crossovers" were performed and progeny equations were generated; (3) the fitness of each progeny equation was assessed by the LOF measure; (4) a progeny equation with better fitness score was preserved. The model with a proper balance of all statistical terms explained the variance of the biological activity. The G/PLS algorithm was used as an alternative to a GFA calculation. The G/PLS is derived from two QSAR calculation methods: GFA and partial least squares (PLS). The G/PLS algorithm uses GFA to select appropriate basis functions to be used in a model of the data and PLS regression as the fitting technique to weigh the basis functions' relative contributions in the final model.

Table 1. Structure of study molecules

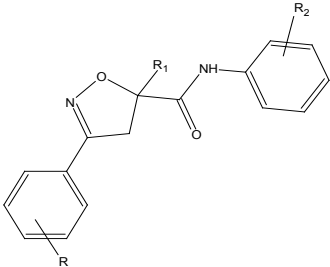
Basic Structure	Molecule Number	R	R ₁	R ₂
	1	<i>m</i> -C=NH(NH ₂)	CH ₂ COOH	<i>p</i> -C=NH(NH ₂)
	2	<i>m</i> -C=NH(NH ₂)	CH ₂ COOCH ₃	<i>p</i> -C=NH(NH ₂)
	3	<i>m</i> -C=NH(NH ₂)	CH ₂ CONHCH ₂ COOCH ₃	<i>p</i> -C=NH(NH ₂)
	4	<i>p</i> -C=NH(NH ₂)	CH ₂ COOCH ₃	<i>m</i> -C=NH(NH ₂)
	5	<i>m</i> -C=NH(NH ₂)	CH ₂ COOCH ₃	<i>m</i> -C=NH(NH ₂)
	6	<i>m</i> -C=NH(NH ₂)	H	<i>p</i> -C=NH(NH ₂)

Table 2. Structure of study molecules

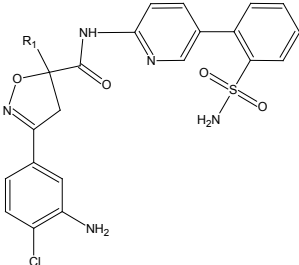



Basic Structure	Molecule Number	R ₁
	7	NHCOCH ₃
	8	NHCOOH
	9	NHSO ₂ NH ₂
	10	NHSO ₂ CH ₃
	11	NHSO ₂ (CH ₂) ₂ CH ₃
	12	NHSO ₂ CH ₂ CF ₃
	13	NHSO ₂ C ₆ H ₅
	14	NHSO ₂ -thiophen-3-yl
	15	NHSO ₂ -3-pyridyl
	16	NHSO ₂ CH ₂ C ₆ H ₅
	17	
	18	
	19	
	20	NHCONH ₂
	21	NHCONHCH ₂ CH ₃

Table 3. Structure of study molecules

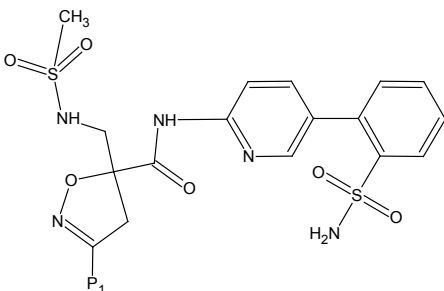
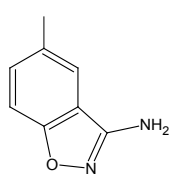
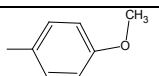
Basic Structure	Molecule Number	P ₁
	22	
	23	

Table 4. Structure of study molecules

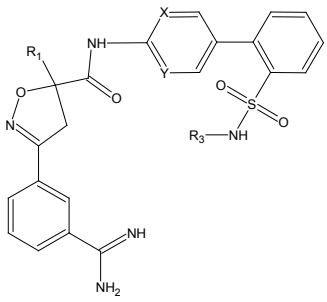
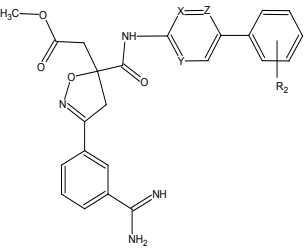
Basic Structure	Molecule Number	R ₁	R ₃	X	Y
	24	CH ₂ CH ₂ OCH ₃	H	CH	CH
	25	COONH ₂	H	CCH ₃	CH
	26	CH ₂ COOH	H	CH	CH
	27	CH ₂ SO ₂ CH ₂ CH ₃	H	CH	CH
	28	CH ₂ OCH ₃	H	CH	CH
	29	CH ₂ OCH ₂ CH ₃	H	CH	CH
	30	CH ₂ O- <i>n</i> -Pr	H	CH	CH
	31	CH ₂ O- <i>i</i> -Pr	H	CH	CH
	32	CH ₂ O- <i>n</i> -Bu	H	CH	CH
	33	CH ₂ O- <i>i</i> -Amyl	H	CH	CH
	34	H	H	CH	CH
	35	CH ₃	H	CH	CH
	36	CF ₃	H	CH	CH
	37	CH ₂ C ₆ H ₅	H	CH	CH
	38	CH ₂ -1-(1,2,4-triazole)	H	CH	CH
	39	CH ₂ -1-tetrazole	H	CH	CH
	40	CH ₂ -2-tetrazole	H	CH	CH
	41	CH ₂ OCH ₃	H	N	N
	42	CH ₂ OCH ₂ CH ₃	H	N	N
	43	CH ₂ SCH ₂ CH ₃	H	N	N
	44	CH ₂ SO ₂ CH ₂ CH ₃	H	N	N
	45	CH ₂ -1-tetrazole	H	N	N
	46	CH ₃	H	N	CH
	47	CH ₂ OCH ₃	H	N	CH
	48	CH ₂ OCH ₂ CH ₃	H	N	CH
	49	CH ₂ SCH ₂ CH ₃	H	N	CH
	50	CH ₂ SO ₂ CH ₂ CH ₃	H	N	CH
	51	CH ₂ -1-tetrazole	H	N	CH
	52	CH ₂ -1-tetrazole	CH ₃	CH	CH
	53	CH ₂ -1-tetrazole	(CH ₂) ₂ CH ₃	CH	CH
	54	CH ₂ -1-tetrazole	H	CCH ₃	CH
	55	CH ₂ -1-tetrazole	H	CF	CH

Table 5. Structure of study molecules

Basic Structure	Molecule Number	R ₂	X	Y	Z
	56	SO ₂ NH ₂	CH	CH	CH
	57	H	CH	CH	CH
	58	3'-CH ₃	CH	CH	CH
	59	2'-CH ₃	CH	CH	CH
	60	3'-CF ₃	CH	CH	CH
	61	2'-CF ₃	CH	CH	CH
	62	3'-OCH ₃	CH	CH	CH
	63	2'-OCH ₃	CH	CH	CH
	64	3'-SO ₂ NH ₂	CH	CH	CH
	65	2'-SH	CH	CH	CH
	66	2'-COOCH ₃	CH	CH	CH
	67	2'-SO ₂ CH ₃	CH	CH	CH
	68	2'-SO ₂ NH ₂	CH	CCH ₃	CH
	69	2'-SO ₂ NH ₂	CH	CF	CH
	70	2'-SO ₂ NH ₂	CH	CH	N
	71	2'-SO ₂ NH ₂	N	CH	CH
	72	2'-SO ₂ NH ₂	N	N	CH

The PLS is a generalization of regression, which can handle data with strongly correlated and/or noisy or numerous X variables. r^2 , r and least square error (LSE) were taken as statistical measures for PLS equations, and LOF was noted for the GFA-derived equations. The 3D-QSAR equations generated were validated by PRESS (leave-one-out) and bootstrap statistics which were calculated using the QSAR+ module of the Cerius2 software and the reported parameters were cross-validation r^2 (q^2), predicted residual sum of squares (PRESS), standard deviation based on PRESS (SPRESS), standard deviation of error of prediction (SDEP) and bootstrap r^2 (bsr²). Additionally, the final models were subjected to leave-20%-out cross validation with 15 trials in each case.

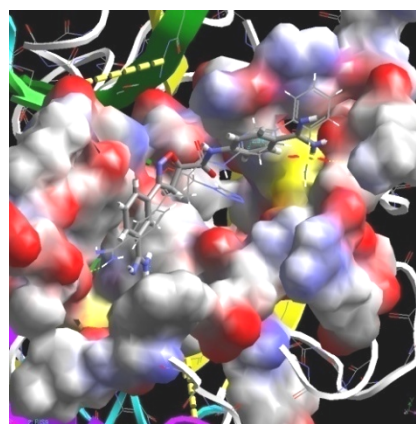
2.3. Pharmacophore Modeling (Catalyst)

All pharmacophore modeling studies were performed using Catalyst 4.11 installed on a Silicon Graphic Octane desktop Workstation. The flexibility of each molecule was represented by a set of energetically reasonable conformers which were generated with the Catalyst catConf module choosing a maximum number of 250 conformers, the *best* quality generation type, and an energy threshold of 20 kcal/mol beyond the calculated global energy minimum. The number of conformers generated for each molecule was limited to a maximum of 255. Ten hypotheses were generated using these conformers for each of the molecules and estimated activity values were generated after selection of the following features for the drugs; hydrogen bond donor, hydrophobic, negative charge, positive ionizable and ring aromatic. After assessing all 10 hypotheses generated for each data set, the lowest energy cost hypothesis was considered the best. The goodness of the structure activity correlation was estimated by means of the correlation coefficient (r). Catalyst also calculates the total energy cost of the generated pharmacophores from the deviation between the estimated activity and the observed activity, combined with the complexity of the hypothesis (i.e., the number of pharmacophore features). A null hypothesis was additionally calculated, which presumes that there is no relationship in the data and that experimental activities are normally

distributed about their mean. Hence, the greater the difference between the energy cost of the generated hypothesis and the energy cost of the null hypothesis, the less likely it is that the hypothesis reflects a chance correlation. This criterion was then used as an assessment of the pharmacophore model selected.

3. Results & Discussion

Results of docking studies were displayed in Table-6. Difference in interaction energy values and the binding mode of conformations (Fig-1) can be used to explain the activity variance between the molecules. Conformer with better interaction energy values for 51 (most active molecule) was isolated from the others and was used as shape reference to align all other study molecules for 3D-QSAR studies.

**Figure 1.** Comparative view on binding modes between high (stick model) and low (cylinder model) active molecules**Table 6.** Dock score and interaction energy values

Molecule Number	Dock Score	Ligand Interaction Energy	Actual Activity (pKi)
51	73.660	-9.682	9.770
3	71.861	-5.479	7.745
14	78.664	-9.339	8.921
17	79.067	-10.071	8.337
56	77.955	-7.561	8.201
60	71.702	-8.440	6.620
25	75.043	-10.075	8.131
71	72.448	-9.925	9.018

Molecular shape analysis

Alignment of molecules for MSA1 and MSA2 study were shown in Fig-2 and 3 respectively. MSA1 study was found to be superior to MSA2. The best equation obtained from GFA regression (at 100,000 crossovers and F value for inclusion of variables was set to 4) in MSA1 is the following:

$$\text{Activity} = -2.1378 - 0.022792 * \langle \text{Foct} \rangle - 0.008514 * \langle \text{NCOSV} \rangle - 0.324377 * \langle \text{AlogP98} \rangle + 0.045833 * \langle \text{Mol Ref} \rangle + 0.69775 * \langle \text{Rad of Gyration} \rangle$$

-----Equation 1, Fig 4

Equation 1 could explain 73.6% of the variance and predict 64.1% of the variance. The final model was also subjected leave-one-out and leave-20%-out cross-validation tests with 15 trials and the r^2 value between the observed and predicted values was found to be 0.727 and 0.743 respectively (Table-8). The following two equations were among those obtained from the G/PLS and Stepwise regressions respectively,

$$\text{Activity} = -3.55282 + 0.06584 * \langle \text{DIPOLE_MOPAC} \rangle - 0.00127 * \langle \text{Vm} \rangle - 1.177066 * \langle \text{AlogP98} \rangle - 0.720516 * \langle \text{HOMO_MOPAC} \rangle + 0.048567 * \langle \text{Mol Ref} \rangle - 0.005897 * \langle \text{NCOSV} \rangle$$

-----Equation 2, Fig 5

$$\text{Activity} = -4.14186 + 0.0064018 * \langle \text{COSV} \rangle + 0.062217 * \langle \text{DIPOLE_MOPAC} \rangle - 0.751026 * \langle \text{HOMO_MOPAC} \rangle + 0.028011 * \langle \text{Mol Ref} \rangle - 0.18584 * \langle \text{AlogP98} \rangle$$

-----Equation 3, Fig 6

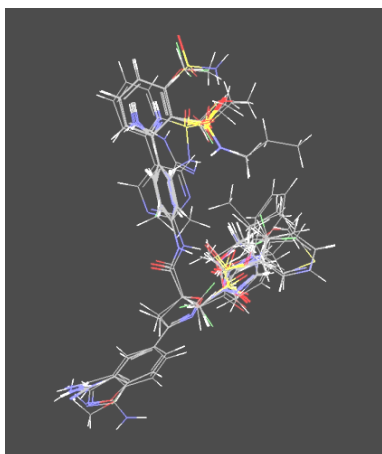


Figure 2. Alignment of molecules for MSA1 study

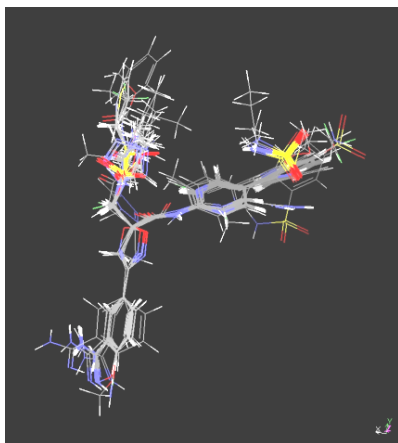


Figure 3. Alignment of molecules for MSA2 study

Table 7. Descriptors Used For MSA Study

Sl. No.	Descriptor Title	Type/Family
1	DIFF V A ³	MSA
2	COS V A ³	MSA
3	F ₀	MSA
4	NCOS V A ³	MSA
5	SHAPE RMS	MSA
6	SR VOL A ³	MSA
7	LUMO_M0Ev	QUANTUM MECHANICAL
8	DIPOLE_Mopac debye	QUANTUM MECHANICAL
9	HF_MOPAC kcal	QUANTUM MECHANICAL
10	HOMO_MOEv	QUANTUM MECHANICAL
11	H-BOND ACCEPTOR	STRUCTURAL
12	H-BOND DONOR	STRUCTURAL
13	ALogP98	THERMODYNAMIC
14	Fh ₂ O	THERMODYNAMIC
15	Foct	THERMODYNAMIC
16	LogP	THERMODYNAMIC
17	MOL REF	THERMODYNAMIC
18	ZAGREB	TOPOLOGICAL
19	RADIUS OF GYRATION	SPATIAL
20	PMI_Mag	SPATIAL
21	SHADOW-XY	SPATIAL
22	SHADOW-XZ	SPATIAL
23	SHADOW-YZ	SPATIAL
24	SHADOW-XY	SPATIAL
25	SHADOW-XZ	SPATIAL
26	SHADOW-YZ	SPATIAL
27	SHADOW-nu	SPATIAL
28	SHADOW-XI	SPATIAL
29	SHADOW-YI	SPATIAL
30	SHADOW-ZI	SPATIAL

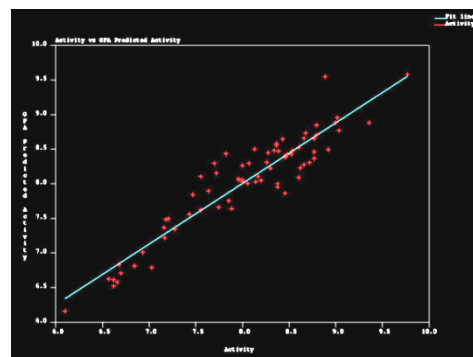


Figure 4. Actual Vs Predicted Activity values obtained in MSA1 by GFA regression

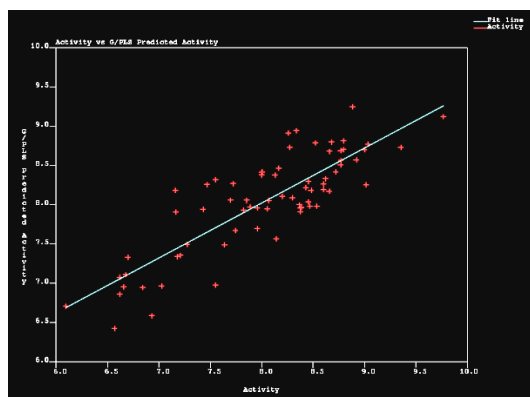


Figure 5. Actual Vs Predicted Activity values obtained in MSA1 by G/PLS regression

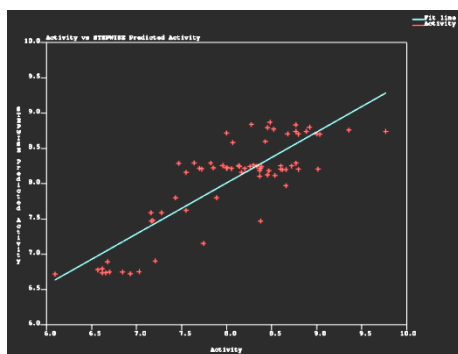


Figure 6. Actual Vs Predicted Activity values obtained in MSA1 by Stepwise regression

Molecular field analysis

The generated field was of the rectangular type. The probes used in the study (MFA1) were H^+ and CH_3 . In another study (MFA2) HO^- probe was also included. The charge method used was Gasteiger and the energy cutoff was kept at -30 to +30 kcal. QSAR equations were generated using both G/PLS and GFA method. The number of iterations was set to 1000,000 to obtain the final equation. The mutation probabilities were set to the system defaults. The final best result was obtained with G/PLS in MFA2 and was discussed here. A view of aligned molecules studied in the field was shown in Fig-7. The following equation was obtained from the MFA:

G/PLS equation

$$\text{Activity} = 6.10935 + 0.101911 H^+/220 + 0.010303 CH_3/674 - 0.01462 H^+/570 + 0.02806 HO^-/569 + 0.015087$$

$$HO^-/653 - 0.039155 H^+/213 + 0.027586 HO^-/659 + 0.017601 H^+/752 + 0.032525 CH_3/870 + 0.013759 H^+/552 - 0.037206 H^+/840 - 0.035787 H^+/527 + 0.002849 H^+/751 - 0.054315 H^+/949$$

----- **Equation 4, Fig 8**

In Eq. 4, $H^+/220$, $CH_3/674$..., and so on were the probes and their numbering (corresponding to spatial positions as shown in Fig-7); i.e., these represent interactions at points 220 by H^+ , 674 by CH_3 , etc. Equation 4 is of very good statistical quality. It shows 90.2% explained variance while leave-one-out cross-validation r^2 is found to be 80.1%. The final models were also subjected to leave-20%-out cross-validation tests with 15 trials and the r^2 value between the observed and predicted values were found to be 0.871 and 0.870 respectively (Table-9). Predicted activity values for each molecule by the QSAR model were given in Table-10. Fig-4, 5, 6 & 8 show the plots between actual and predicted activity values.

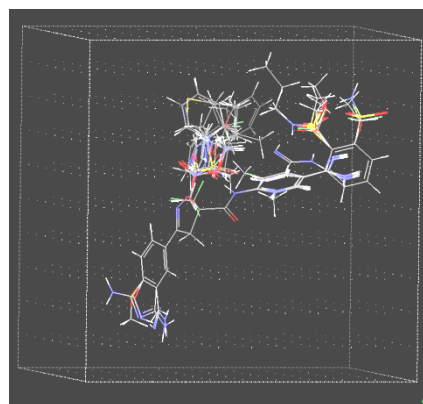


Figure 7. Alignment of molecules for MFA study

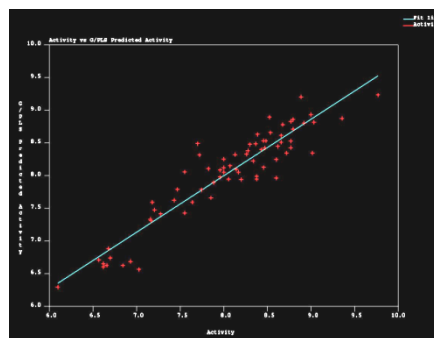


Figure 8. Actual Vs Predicted Activity values obtained in MFA2 by G/PLS regression

Table 8. Statistics of MSA QSAR Equation

Statistical Parameters	MSA1			MSA2		
	G/PLS	GFA	STEPWISE	G/PLS	GFA	STEPWISE
r^2	0.730	0.736	0.731	0.695	0.696	0.677
q^2	0.639	0.641	0.623	0.648	0.628	0.605
LOO r^2	0.727	0.727	0.721	0.682	0.686	0.648
20% out r^2	0.780	0.743	0.691	0.694	0.753	0.724
BS r^2	0.695	0.736	0.731	0.679	0.696	0.677
F-Test		30.211	35.841		30.241	35.136
PRESS	14.878	14.733	15.516	14.516	15.347	
LOF		0.217			0.234	
LSE	0.155	0.151		0.175	0.174	
Outliers	3	7	7	1	4	5

Table 9. Statistics of MFA QSAR Equation

Statistical Parameters	MFA1		MFA2	
	G/PLS	GFA	G/PLS	GFA
r^2	0.881	0.902	0.893	0.877
q^2	0.666	0.816	0.715	0.840
LOO r^2	0.777	0.801	0.807	0.829
20% out r^2	0.870	0.871	0.871	0.868
BS r^2	0.740	0.903	0.797	0.877
F-Test		37.562		49.226
PRESS	13.780	7.568	11.742	6.608
LOF		0.150		0.125
LSE	0.068	0.056	0.061	0.070
Outliers	4	8	4	6

Table 10. QSAR Activity Table

Molecule Number	Actual Activity (pK _i)	MFA ²		MSA ¹		Stepwise Predicted
		G/PLS Redicted	GFA Predicted	G/PLS Predicted	GFA Predicted	
1	6.845	6.625	6.813	6.990	6.945	6.941
2	7.027	6.568	6.788	6.990	7.032	7.032
3	7.745	7.785	7.658	7.457	7.711	7.780
4	6.932	6.686	7.008	6.962	7.027	7.032
5	6.097	6.294	6.159	6.997	7.053	7.043
6	6.569	6.710	6.622	6.674	6.529	6.438
7	7.959	7.981	8.068	7.766	7.794	7.802
8	7.824	8.108	8.433	7.722	7.675	7.701
9	8.260	8.333	8.314	8.407	8.447	8.466
10	8.721	8.347	8.304	8.331	8.293	8.310
11	8.432	8.400	8.646	8.386	8.326	8.354
12	8.481	8.533	8.425	8.200	8.206	8.308
13	8.770	8.829	8.367	8.629	8.565	8.548
14	8.921	8.807	8.495	8.613	8.587	8.580
15	8.523	8.899	8.427	8.553	8.702	8.714
16	8.658	8.615	8.658	8.617	8.593	8.614
17	8.337	8.223	8.480	8.659	8.782	8.793
18	8.000	8.124	8.032	8.463	8.468	8.552
19	8.000	8.255	8.264	8.642	8.638	8.494
20	7.854	7.664	7.754	8.133	7.970	7.897
21	7.468	7.789	7.842	8.409	8.285	8.187
22	8.602	7.966	8.090	7.661	7.750	7.747
23	7.161	7.335	7.368	7.716	7.859	7.869
24	8.300	8.481	8.224	8.007	8.115	8.130
25	8.131	8.326	8.499	8.246	8.462	8.478
26	7.699	8.492	8.295	8.121	8.121	8.144
27	8.456	8.127	7.863	8.471	8.515	8.593
28	8.469	8.422	8.416	8.161	8.129	8.124
29	8.456	8.535	8.385	8.225	8.156	8.167
30	8.538	8.657	8.476	8.337	8.164	8.190
31	8.367	8.789	8.554	8.319	8.222	8.252
32	7.886	7.895	7.644	8.483	8.181	8.220
33	7.432	7.623	7.558	8.376	8.225	8.308
34	8.143	8.101	8.031	7.907	7.772	7.693
35	7.959	8.087	8.062	8.003	7.841	7.805
36	7.638	7.595	7.894	7.855	7.720	7.766
37	8.070	8.150	8.294	8.480	8.253	8.198
38	8.770	8.532	8.652	8.594	8.617	8.556
39	8.796	8.862	8.850	8.860	8.879	8.768
40	8.796	8.714	8.709	8.733	8.679	8.579
41	8.004	8.053	8.055	8.050	8.101	8.121
42	8.168	8.055	8.106	8.167	8.158	8.186
43	8.620	8.448	8.229	8.279	8.166	8.126
44	8.276	8.381	8.448	8.498	8.613	8.679
45	8.886	9.209	9.550	8.723	8.887	8.836
46	8.387	8.633	8.473	7.885	7.694	7.682
47	8.602	8.248	8.526	7.931	7.951	7.955
48	7.553	8.057	8.104	8.178	8.015	8.041
49	8.658	8.514	8.280	8.220	7.985	7.947
50	8.770	8.432	8.461	8.538	8.481	8.572
51	9.770	9.237	9.584	9.819	9.844	9.863
52	9.036	8.819	8.769	8.798	8.920	8.832
53	8.678	8.783	8.734	9.003	8.981	8.944
54	9.000	8.937	8.888	8.790	8.908	8.801
55	9.357	8.877	8.886	8.711	8.853	8.761
56	8.201	7.943	8.046	7.916	8.039	8.073
57	6.657	6.629	6.577	6.693	6.567	6.646
58	6.620	6.653	6.611	6.991	6.926	6.874
59	6.678	6.882	6.829	6.859	6.893	6.840
60	6.620	6.606	6.521	7.152	7.037	7.090

61	7.553	7.430	7.621	7.070	7.024	7.069
62	6.699	6.742	6.706	6.982	7.036	7.037
63	7.208	7.479	7.493	6.806	6.966	6.953
64	7.168	7.308	7.219	7.800	7.968	8.013
65	7.180	7.595	7.482	7.000	7.076	6.996
66	7.276	7.416	7.344	7.109	7.243	7.258
67	8.377	7.990	7.998	8.167	8.085	8.121
68	8.056	7.948	8.009	8.184	8.261	8.285
69	8.377	7.950	7.952	7.895	8.089	8.138
70	7.721	8.317	8.154	8.171	8.204	8.233
71	9.018	8.348	8.959	8.212	8.147	8.213
72	8.367	8.739	8.585	8.034	8.211	8.267

Pharmacophore Model

Catalyst uses a collection of molecules with activity spanning orders of magnitude to construct a useful model of the chemical features and their position in 3D space (Fig-9 and Table-11) necessary for a biological response. After several iterations the model of our 30-molecule data set produced a good correlation when compared with the estimated (r of 0.92, Table-12). The model contained four features necessary for activity, namely two hydrogen bond donor, a hydrophobic aromatic, and a ring aromatic feature (Fig-10 and Table-11). The features were compared with highest and low activity molecules in the dataset using comparefit and found that ring aromatic feature was missed for the later one (Fig-10 & 11). The generated catalyst hypotheses can serve as query features in 3D database (DB) search for virtual screening to detect novel lead compounds as factor Xa inhibitors.

Table 11. Pharmacophore Model

DISTANCE	Ra-Hd ₂	Ra-Hd ₁	Ra-La	La-Hd ₂	La-Hd ₁	Hd ₁ -Hd ₂
Min.	5.800	4.666	4.153	9.810	4.338	10.891
Max.	7.800	6.666	6.153	11.810	6.338	12.891
ANGLE	Hd ₁ RaHd ₂	Hd ₁ LaHd ₂	Hd ₁ RaLa	LaRaHd ₂		
Min.	139.0 °	83.1 °	53.9 °	124.0 °		
Max.	149.0 °	93.0 °	63.9 °	133.9 °		

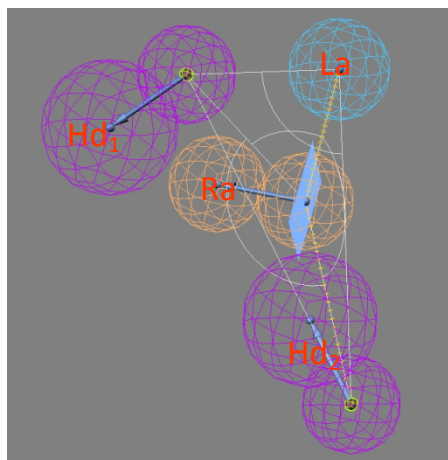
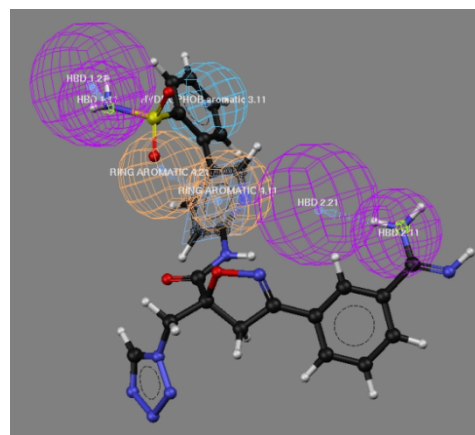


Figure 9. Pharmacophore model showing various features required for activity. (Ra- Ring aromatic; La- Hydrophobic aromatic; Hd- Hydrogen bond donor)

Table 12. Pharmacophore Activity Table

Molecule Number	Fit	Actual Activity Ki(Nm)	Estimated Activity Ki(Nm)	Error
51	8.66	0.17	0.16	-1.1
71	7.57	0.96	2	+2.0
50	7.61	1.7	1.8	+1.1
14	7.19	2.8	4.7	+1.7
11	6.70	3.7	15	+4.0
17	6.61	4.6	18	+3.9
9	7.14	5.5	5.2	-1.1
56	7.26	6.3	4.0	-1.6
25	6.92	7.4	8.9	+1.2
68	7.29	8.8	3.8	-2.3
41	6.74	9.9	13	+1.3
7	6.67	11	16	+1.4
32	6.87	13	10	-1.3
20	6.73	14	14	-1
8	6.68	15	15	+1
3	5.98	18	77	-1.6
70	6.78	19	12	-1.3
26	6.67	20	16	-1.4
36	6.65	23	16	+1.1
61	6.38	28	30	-2.4
21	6.71	34	14	-1
66	6.15	53	52	-1.7
64	6.26	68	40	-1.3
2	6.01	94	72	-1.3
4	5.90	120	92	-1.3
1	5.96	140	81	-1.8
62	5.68	200	150	-1.3
59	5.87	210	98	-2.1
58	5.62	240	180	-1.4
6	5.33	270	340	+1.3



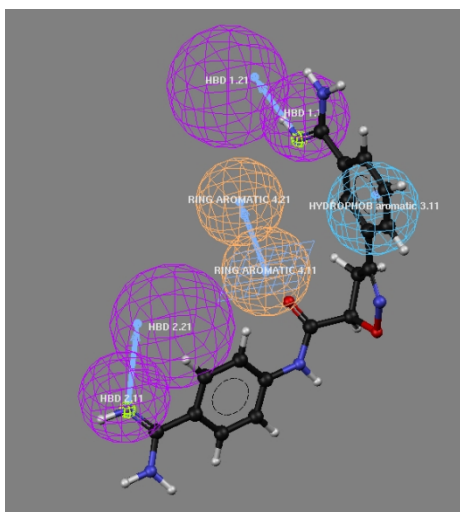


Figure 11. Partial alignment of Pharmacophoric features observed for the low active compound of the series

4. Conclusions

The present 3D-QSAR analysis based on the alignment of best docked conformation explores the spatial, shape and thermodynamic requirements for the binding affinity of aryl amidino isoxazoline derivatives to factor Xa. The MSA-derived equations shows the importance of thermodynamic and quantum mechanical descriptors, molecular refractivity and radius of gyration contribution to activity. The MFA-derived equation shows interaction energies at different grid points with positive, negative and neutral bulk probes. Statistically reliable 3D-QSAR and pharmacophore models obtained from this study on aryl amidino Isoxazoline derivatives suggest that these techniques could be useful to design potent factor Xa inhibitors.

ACKNOWLEDGEMENTS

Authors are thankful to GVK biocampus, Hyderabad, India for the facilities and technical support. The corresponding author thank the Acharya Nagarjuna University, Guntur, Andhrapradesh, India for the support of Part-time Ph.D in Pharmacy.

REFERENCES

- [1] Edgar, F.M.; Stanley, M.S.; Jocelyn, A.W. Molecular modelling and drug design *Pharm. & Ther.* 2000, 85(3), 113–121.
- [2] Diane, J.M. Computational approaches to structure-based ligand design. *Pharm. & Ther.* 1999, 84(2), 179–191.
- [3] Christoph, S.; Gerhard, K. Identification and mapping of small-molecule binding sites in proteins: computational tools for structure-based drug design. *II Farmaco*, 2002, 57(3), 243–251
- [4] Venkatachalam, C.M.; Jiang, X.; Oldfield, T.; Waldman, M. LigandFit: A Novel Method For The Shape-Directed Rapid Docking Of Ligands To Protein Active Sites. *J Mol Graph Model.* 2003 21(4), 289-307.
- [5] Suresh, B.V.; Niles, R.T.; Mariam, S.D. Quantitative Structure-Activity Relationship Study of Some Aspartic Acid Analogues to Correlate and Predict their Sweetness Potency. *QSAR Comb. Sci.* 2007, 26(2), 204–214
- [6] Charifson P.S. Practical applications of computer-aided drug design' 1st ed.; Marcel Dekker Inc., New York, 1997; pp. 105–163.
- [7] Barnum, D.; Greene, J.; Smellie, A.; Sprague, P. Pharmacophore perception, development, and use in drug design *J. Chem. Inf. Comput. Sci.*, 1996, 36, 563-571.
- [8] Schuster, D.; Maurer, E.M.; Laggner, C.; Nashev, L.G.; Wilckens, T.; Langer T.; Odermatt, A. The discovery of new 11 β -hydroxysteroid dehydrogenase type 1 inhibitors by common feature pharmacophore modeling and virtual screening. *J. Med. Chem.*, 2006, 49(12), 3454-66.
- [9] Maduskuie, T.P.; Jr, McNamara K.J.; Ru, Y.; Knabb, R.M.; Stouten, P.F. Rational design and synthesis of novel, potent bis-phenylamidine carboxylate factor Xa inhibitors. *J. Med. Chem.* 1998, 41(1), 53-62
- [10] Matter, H.; Defossa, E.; Heinelt, U.; Blohm, P.M.; Schneider, D.; Mueller, A.; Herok, S.; Schreuder, H.A.; Liesum, A.; Brachvogel, V.; Loenze, P.; Walser, A.; Al-Obeidi, F.; Wildgoose, P. Design and quantitative structure-activity relationship of 3-amidinobenzyl-1H-indole-2-carboxamides as potent, nonchiral, and selective inhibitors of blood coagulation factor Xa. *J. Med. Chem.* 2002, 45, 2749-2769
- [11] Pancras C.W.; Mimi, L.Q.; Earl, J.; Crain, C.A.; Watson, R.W.; Robert, M.K. Nonpeptide Factor Xa Inhibitors: I. Studies with SF303 and SK549, a New Class of Potent Antithrombotics. *J. Pharm. Exp. Ther.* 2000, 292, 351–357.
- [12] Quan, M. L.; Pruitt, J.R.; Ellis, C.D.; Liauw, A.Y.; Galemno, Jr.R.A.; Stouten, P.F.W.; Wityak, J.; Knabb, R.M.; Thoolen, M.J.; Wong, P.C.; Wexler, R.R. Nonbenzamidine isoxazoline derivatives as factor Xa inhibitors. *Bio. & Med. Chem. Lett.*, 2003, 13, 1023–1028
- [13] Quan, M.L.; Liauw, A.Y.; Ellis, C.D.; Pruitt, J.R. Carini, D.J.; Bostrom, L.L.; Huang, P.P.; Harrison, K.; Knabb, R.M.; Thoolen, M.J.; Wong, P.C.; Wexler, R.R. Design and synthesis of isoxazoline derivatives as factor Xa inhibitors. 1. *J. Med. Chem.* 1999, 42, 2752-2759.
- [14] Quan, M. L.; Pruitt, J.R.; Ellis, C.D.; Liauw, A.Y.; Galemno, Jr.R.A.; Stouten, P.F.W.; Wityak, J.; Knabb, R.M.; Thoolen, M.J.; Wong, P.C.; Wexler, R.R. Bisbenzamidine isoxazoline derivatives as factor Xa inhibitors. *Bio. & Med. Chem. Lett.*, 1997, 7(21), 2813-2818.
- [15] Quan, M.L.; Liauw, A.Y.; Ellis, C.D.; Pruitt, J.R.; Carini, D.J.; Bostrom, L.L.; Huang, P.P.; Harrison, K.; Knabb, R.M.; Thoolen, M.J.; Wong, P.C.; Wexler, R.R. Design and synthesis of isoxazoline derivatives as factor Xa inhibitors. 2. *J. Med. Chem.* 1999, 42, 2760-2773.
- [16] Cerius2 Version 4.11 (2006) Accelrys Inc, San Diego, USA; <http://www.accelrys.com/cerius2>