

Original Article

2D AND 3D-QSAR ANALYSIS OF AMINO (3-((3, 5-DIFLUORO-4-METHYL-6-PHENOXYPYRIDINE-2-YL) OXY) PHENYL) METHANIMINIUM DERIVATIVES AS FACTOR Xa INHIBITOR

SMITA SUHANE¹, A. G. NERKAR², KUMUD MODI³, SANJAY D. SAWANT⁴

¹Department of Applied Science, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune 411018 India, ²Department of Pharmaceutical Chemistry, R. C. Patel Institute of Pharmaceutical Education and Research, Karwand Naka, Shirpur 425405 India, ³P. M. B Gujarati Science College, 1, Nasia Road, Indore 452001 India, ⁴Department of Pharmaceutical Chemistry, Smt. Kashibai Navale College of Pharmacy, Kondhwa-Saswad Road, Kondhwa (Bk.), Pune 411048
Email: smita3472@gmail.com

Received: 30 Jun 2017 Revised and Accepted: 08 Jan 2019

ABSTRACT

Objective: The main objective of the present study was to evolve a novel pharmacophore of methaniminium derivatives as factor Xa inhibitors by developing best 2D and 3D QSAR models.

The models were developed for amino (3-((3, 5-difluoro-4-methyl-6-phenoxy pyridine-2-yl) oxy) phenyl) methaniminium derivatives as factor Xa inhibitors.

Methods: With the help of Marvin application, 2D structures of thirty compounds of methaniminium derivatives were drawn and consequently converted to 3D structures. 2D QSAR using multiple linear regression (MLR) analysis and PLS regression method was performed with the help of molecular design suite VLife MDS 4.3.3. 3D QSAR analysis was carried out using k-Nearest Neighbour Molecular Field Analysis (k-NN-MFA).

Results: The most significant 2D models of methaniminium derivatives calculated squared correlation coefficient value 0.8002 using multiple linear regression (MLR) analysis. Partial Least Square (PLS) regression method was also employed. The results of both the methods were compared. In 2D QSAR model, T_C_0_5, T_2_0_2, s log p, T_2_0_1 and T_2_0_6 descriptors were found significant.

The best 3D QSAR model with k-Nearest Neighbour Molecular Field Analysis have predicted q^2 value 0.8790, q^2_{se} value 0.0794, $pred\ r^2$ value 0.9340 and $pred\ r^2_{se}$ value 0.0540. The stepwise regression method was employed for anticipating the inhibitory activity of this class of compound. The 3D model demonstrated that hydrophobic, electrostatic and steric descriptors exhibit a crucial role in determining the inhibitory activity of this class of compounds.

Conclusion: The developed 2D and 3D QSAR models have shown good r^2 and q^2 values of 0.8002 and 0.8790 respectively. There is high agreement in inhibitory properties of experimental and predicted values, which suggests that derived QSAR models have good predicting properties.

The contour plots of 3D QSAR (k-NN-MFA) method furnish additional information on the relationship between the structure of the compound and their inhibitory activities which can be employed to construct newer potent factor Xa inhibitors.

Keywords: QSAR, K-Nearest Neighbour Molecular Field Analysis, Amino (3-((3, 5-difluoro-4-methyl-6-phenoxy pyridine-2-yl) oxy) phenyl) methaniminium derivatives, Factor Xa

© 2019 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)
DOI: <http://dx.doi.org/10.22159/ijpps.2019v11i2.21067>

INTRODUCTION

The structure of any molecule dictates its properties. By modifying the chemical structure of any compound, its biological activities also get changed. In other words, the biological activity of a compound is a function of its chemical structure. QSAR suggests that if a group of chemicals show the same mechanism of action towards a target then alteration in the biological activity also alters chemical, structural and physical properties [1].

QSAR methods are not only used in drug designing but are widely used in other sciences too, i.e. in biology, toxicology [2-3] environmental toxicology [4], agrochemistry, pharmaceutical chemistry etc. The QSAR is also used to determine the initial and final point of synthesis [5]. This, in turn, reduces the number of compounds that could be practically/experimentally synthesized. This part of QSAR is not only beneficial for the pharmaceutical industry but also to environmental regulatory authority and human beings for the reduction in toxic effects [6-10].

QSAR models are not only used for prediction of properties but are also helpful in selection of alternative mechanism of action, determination of useful structural characteristics, projecting new

design methodologies and help in proposing new hypotheses for future research work [11]. Thus, QSAR decreases the cost, time and human resources to make that drug reachable to the patient.

The injury to the blood vessel causes the body to use platelets and fibrin to make a blood clot or thrombus. The process of formation of the blood clot is called thrombosis or coagulation. When clotting takes place excessively and the thrombus or clot breaks and detaches from its side then an embolus is produced [12-13]. When thrombus occlude the blood vessel then the supply of oxygen to the tissue through blood is interrupted which causes cell death or necrosis.

The thrombi in the arteries are platelet rich with small fibrin requiring antiplatelet therapy primarily in ischemic heart disease. For venous thromboembolism and atrial thromboembolism having large fibrin burden, for example, deep vein thrombosis (DVT), pulmonary embolism (PE), stroke prevention in atrial fibrillation (AF) [14-16] and mechanical prosthetic heart valve, long term anticoagulation is necessary.

Venous arterial thromboembolism continues to be a major cause of morbidity and mortality globally. Research is going on for long term

prevention of thromboembolic episode because of the large burden of other thrombotic and thromboembolic disorders.

Unfractionated heparin was the first anticoagulant (injectable) developed a century ago. Later on, heparin was further modified and refined resulting in the development of low molecular weight heparin and site-specific (injectable) fondaparinux (factor Xa indirect inhibitor).

Similarly, in the field of oral anticoagulants, classical vitamin K antagonists have developed around 70 y ago. Vitamin K antagonist acts by competitively inhibiting enzymes involved in the synthesis of factor II, VII, IX and X in the liver. Around thirty proteins are needed for coagulation cascade [17]. Vitamin K antagonist are having many limitations such as the need for INR monitoring due to the narrow therapeutic window, numerous drug and food interactions and increased chances of bleeding. Therefore, the need for site-specific oral anticoagulant having less chance of bleeding, less drug interaction and not requiring monitoring was sought and development of other antithrombotic agents started [18-23].

At present, heparin and oral anticoagulants are only available therapy for treating thrombotic disorders. Previously, benzamidine

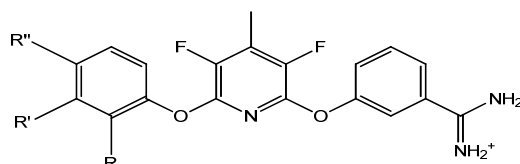
[24] and arylamide compounds have been reported as factor Xa inhibitors [25] but still, oral anticoagulants which do not require monitoring and have less chances of bleeding are essayed to decrease the mortality and morbidity rate.

Factor Xa is present at the central position of coagulation cascade or at the converging point of the extrinsic and intrinsic pathway. The intrinsic and extrinsic pathway model splits up the coagulation process into two fragments. Consequently, Factor Xa has emerged as an attractive target for the development of antithrombotic drugs.

MATERIALS AND METHODS

Software which was employed to present study includes Marvin Sketch, Chem Bio Draw Ultra 12.0 and Vlife MDS 4.3.3 [26]. In the present study, a dissimilar set of amino (3-((3, 5-difluoro-4-methyl-6-phenoxy pyridine-2-yl) oxy) phenyl) methaniminium derivatives were assessed as factor Xa inhibitor as antithrombotic agents. Set of 30 molecules was used as a dataset with their inhibitory activities towards factor Xa. The molecular structures of amino (3-((3, 5-difluoro-4-methyl-6-phenoxy pyridine-2-yl) oxy) phenyl) methaniminium derivatives with their binding affinity for factor Xa are presented in table 1.

Table 1: The structures of compound and Ki values for factor Xa



Compound	R'	R''	R	Ki
3		H	H	5.920
6		H	H	5.886
8		H	H	5.869
10			H	5.850
13		H	H	5.769
15		H	H	5.744
16		H	H	5.744
18			H	5.698
21		H	H	5.677
23		H	H	5.657
26		H		5.619
27		H		5.619
32		H	H	5.494
33		H	H	5.481
36		H		5.346


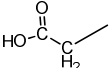






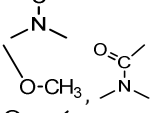










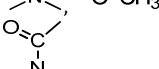



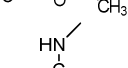
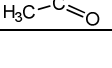
39		H	—CH ₃	5.301
40		H	H	5.301
41		H	H	5.301
42		H		5.300
43		H	H	5.301
44		H		5.301
45		H		5.301
46			H	5.301
47				5.301
48		H		5.301
49		H	H	5.301
50		H	H	5.301
51		H		5.301
52		H		5.301
53		H	H	5.283
55		H	H	5.236

Table 2-i: Molecular descriptors of (Model 1 MLR) training set used in the regression analysis

Compound	T_2_0_1	T_C_0_5	T_2_0_2	T_0_0_6	Slogp
3	6	6	11	0	4.247
6	4	2	8	0	5.19
10	5	4	10	0	5.435
13	6	5	11	0	3.944
16	5	6	10	1	3.334
21	6	8	11	0	4.301
23	4	2	8	0	4.676
26	4	2	8	0	4.119
27	6	6	11	0	3.944
32	5	3	9	0	4.547
33	4	4	8	0	4.911
36	6	7	9	2	4.164
40	6	4	11	0	4.235
41	6	5	11	1	4.247
43	7	7	13	1	4.256
44	6	4	11	0	3.944
45	7	6	13	1	3.65
46	6	5	11	1	3.944
47	5	3	10	0	4.242
49	7	6	13	0	3.953
52	5	5	8	1	4.495
53	4	3	9	0	3.877
55	5	4	10	0	4.545

RESULTS

The most remarkable 2D QSAR models employing multiple linear regression method (MLR) evaluated squared correlation coefficient value 0.8002. The result of MLR analysis were compared with the results of Partial Least Square (PLS) regression method. T_C_O_5, T_2_O_2, s log p, T_2_O_1 and T_2_O_6 descriptors were found significant in 2 D QSAR model.

k-Nearest Neighbour Molecular Field Analysis of best 3D QSAR model predicted q²value 0.8790, q_{2-se} value 0.0794, pred r² value 0.9340 and pred_r²se value 0.0540. For predicting the inhibitory activity of methaniminium derivatives, stepwise regression method was used. The 3D model revealed that hydrophobic, electrostatic and steric descriptors exhibit a critical role in determining the inhibitory activity of this class of compounds.

DISCUSSION

2D-QSAR model

Dataset of 30 compounds was taken into consideration along with their inhibitory activities. 2D-QSAR analysis was performed on the reported compounds of the training set (23) and test set (7). The model obtained thus, observed to be statistically significant. Therefore, the training set (23) and test set (7) were evaluated for the similarity of the distribution patterns of the molecules. Results indicated that the maximum of the test

set is lesser than the maximum of the train set and the minimum of the test set is greater than that of the train set, which is required for further QSAR study.

The minimum 'inhibitory activity' of the test set was greater than the minimum activity of the training set and the maximum activity of the test set was less than the maximum activity of the training set. This indicated that the test set was within the activity domain of the training set. Higher mean value of test set than the training set indicated the presence of relatively more potent compounds in the test set as compared to inactive ones. Table 2-i, ii, iii and iv represent various descriptors used in training and test set of MLR and PLS regression analysis.

Best models using multiple linear regression and partial least square method

Model-1 (MLR)

$$K_i = 0.1397(\pm 0.0209) (T_C_O_5) + 0.2246(\pm 0.0270)(T_2_O_2) + 0.1146 (\pm 0.0440) (s \log p)$$

$$-0.4670(\pm 0.0514) (T_2_O_1) - 0.0994(\pm 0.0334) (T_2_O_6) + 5.2416$$

Model-2 (PLS)

$$K_i = 0.1411(T_C_Cl_4) + 0.2032(T_N_O_5) - 0.0398 (SsOHcount) + 1.1804(chiV3Cluster) + 0.0658(T_T_O_5) - 0.6267(SdOcount) + 4.7066$$

Table 2-ii: Molecular descriptors of (Model 1 MLR) test set used in the regression analysis

Compound	T_2_O_1	T_C_O_5	T_2_O_2	T_O_O_6	slogp
8	4	3	8	0	6.28
15	5	4	9	0	5.376
18	5	5	9	0	5.657
39	5	3	9	0	3.924
48	4	2	8	0	4.537
50	6	4	11	0	3.944
51	6	4	11	0	4.714

Table 2-iii: Molecular descriptors of (Model 2 PLS) training set used in the regression analysis

Compound	T_C_Cl_4	T_N_O_5	SsOHcount	chiV3Cluster	T_T_O_5	SdOcount
10	0	3	0	1.003	15	1
13	1	2	0	0.791	6	0
15	0	2	0	0.769	7	0
16	0	2	0	0.67	8	0
21	0	2	0	1.033	10	1
23	0	2	1	1.039	12	1
27	0	2	1	1.009	12	1
32	0	2	0	0.665	6	0
33	0	2	0	0.699	6	0
39	0	3	0	1.09	10	1
3	0	2	0	1.036	8	0
41	0	4	0	1.067	11	1
43	0	2	1	0.731	12	1
45	0	5	0	1.056	18	1
46	0	4	1	1.029	14	1
47	0	5	2	1.066	17	1
49	0	2	1	0.677	7	0
50	0	2	0	0.602	6	0
51	0	6	1	1.059	18	1
52	0	4	1	0.991	12	1
53	0	2	0	0.727	12	1
55	0	2	0	0.788	9	1
6	0	2	1	0.72	9	0

Table 2-iv: Molecular descriptors of (Model 2 PLS) test set used in the regression analysis

Compound	T_C_Cl_4	T_N_O_5	SsOHcount	chiV3Cluster	T_T_O_5	SdOcount
8	0	2	0	0.67	8	0
18	0	3	1	1.008	14	1
26	0	3	0	1.007	12	1
36	0	4	1	0.991	14	1
42	0	2	1	0.785	11	1
44	0	4	0	1.005	13	1
48	0	4	1	1.01	13	1

Table 3: Statistical parameters (uni-column statistics of model 1 MLR and model 2 PLS) for activity distribution in training and test sets

Parameters	Model 1 MLR		Model 2 PLS	
	Training set	Test set	Training set	Test set
Max.	5.9200	5.8600	5.9200	5.8600
Min.	5.2300	5.3000	5.2300	5.3000
Std. Dev.	0.2304	0.2527	0.2364	0.2317
Sum	126.2200	38.4900	126.3100	38.4000
Average	5.4878	5.4986	5.4917	5.4857

Table 4: It shows the summary of statistical parameters for 2D-QSAR models of Amino (3-((3, 5-difluoro-4-methyl-6-phenoxy pyridine-2-yl) oxy) phenyl) methaniminium derivative

Statistical parameter	Best model (MLR)	Best model (PLS)
N	23	23
Degree of freedom	17	18
r ²	0.8002	0.7813
q ²	0.6107	0.5762
F test	13.6181	16.0749
r ² _{se}	0.1172	0.1222
q ² _{se}	0.1636	0.1702
pred_r ²	0.5248	0.6621
pred_r ² _{se}	0.1744	0.1347

MLR analysis with these newly formed datasets showed better statistically significant results (table 3 and 4). In the multiple regression method, T_C_0_5, T_2_0_2,s log p,T_2_0_1 and T_2_0_6 were found contributing as descriptors while in Partial Least square method, T_C_Cl_4,T_N_0_5,SsOHcount,chiV3Cluster,T_T_0_5 and SdOcount were used as descriptors. On the basis of statistical analysis, the compounds have depicted residual value (table 5) less than 1 unit.

Table 5: It represents residuals of 2 D QSAR

Compound No.	-log (Ki)for factor Xa Actual	Model 1 MLR		Model 2 PLS	
		Predicted Ki	Residual	Predicted Ki	Residual
3	5.9208	5.6385	0.2824	6.0492	0.2093
6	5.8861	5.6468	0.2393	5.7021	0.0004
8	5.8697	5.7126	0.1571	5.6171	0.0718
10	5.8500	5.7376	0.1124	5.6407	0.1229
13	5.7696	5.4640	0.3055	5.7696	0.1490
15	5.7447	5.6057	0.1390	5.6682	0.1195
16	5.7447	5.6768	0.0679	5.6171	0.0108
18	5.6990	5.5788	0.1202	5.5410	0.1619
21	5.6778	5.7252	-0.0474	5.5505	0.0062
23	5.6576	5.5879	0.0697	5.6492	0.0103
26	5.6198	5.5240	0.0958	5.4481	-0.0399
27	5.6198	5.5043	0.1155	5.6138	0.0223
32	5.4949	5.3710	0.1239	5.4797	-0.1146
33	5.4815	5.8942	-0.4128	5.5199	-0.1292
36	5.3468	5.2201	0.1267	5.3177	0.0164
39	5.3010	5.2996	0.0015	5.4146	0.0500
40	5.3010	5.3576	-0.0566	5.2836	0.0144
41	5.3010	5.2999	0.0012	5.2500	-0.0084
43	5.3010	5.3636	-0.0626	5.2856	-0.1942
44	5.3010	5.2249	0.0762	5.3084	-0.0626
45	5.3010	5.1545	0.1466	5.4942	-0.0605
46	5.3010	5.2651	0.0359	5.3626	0.0256
47	5.3010	5.4611	-0.1601	5.3605	-0.2198
48	5.3010	5.5719	-0.2709	5.2744	-0.1054
49	5.3010	5.2886	0.0124	5.5198	0.0453
50	5.3010	5.2249	0.0762	5.4054	0.1138
51	5.3010	5.4125	-0.1115	5.2547	-0.0608
52	5.3010	5.3205	-0.0194	5.1862	0.0345
53	5.2840	5.7611	-0.4771	5.3208	0.1779
55	5.2366	5.6356	-0.3990	5.1955	0.2429

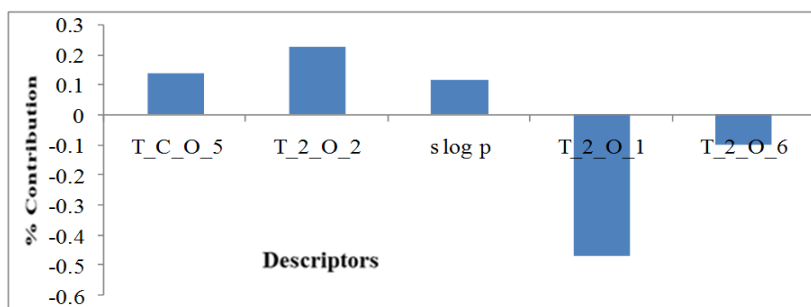


Fig. 1: It shows percentage contribution of descriptors of 2D QSAR (model 1 MLR)

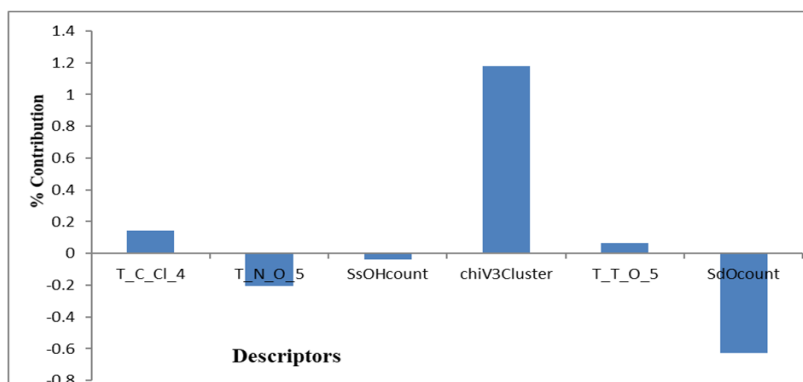


Fig. 2: It shows percentage contribution of descriptors of 2D QSAR (model 2 PLS)

The 'fitness plot' which is a plot of experimental activities with predicted activity of test set compounds and training set compounds for each model shows that built models are

statistically significant. This provides an idea about the fitness of the model and also about the predictive ability of external test set (fig. 3 and fig. 4).

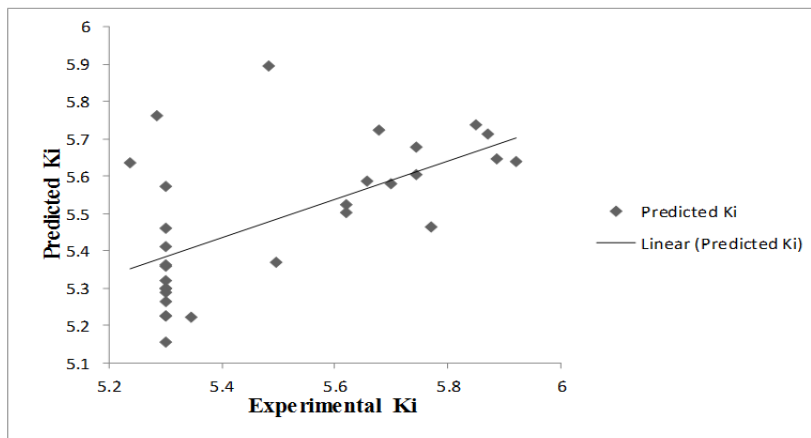


Fig. 3: It shows experimental vs. predicted activities for training and test set molecules from the best predictive MLR model (model 1)

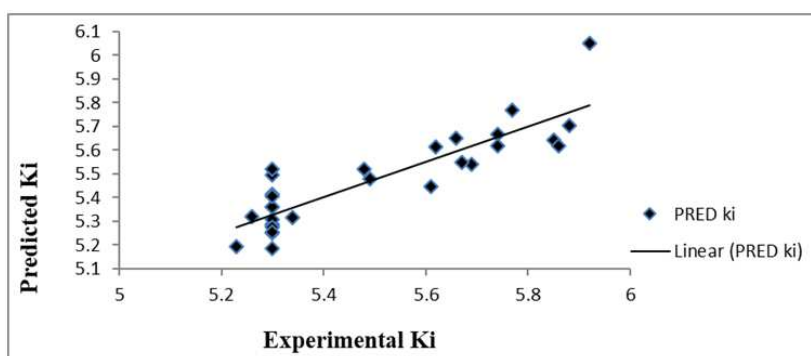


Fig. 4: It represents experimental vs. predicted activities for training and test set molecules from the best predictive PLS model (model 2)

In the generated model of QSAR by MLR method, physicochemical descriptor *slogp* and alignment independent (AI) descriptors *T_C_O_5* and *T_2_O_2* contributed positively, therefore, molecules showing higher values of *T_C_O_5*, *T_2_O_2* and *slogp* will have good inhibitory activity.

The model indicates the negative contribution of alignment independent topological descriptors *T_2_O_1* and *T_2_O_6* thus molecules showing higher values of *T_2_O_1* and *T_2_O_6* will show reduced inhibitory activity. Physicochemical descriptor *slogp* signifies log of the octanol/water partition coefficient. Molecules with high partition coefficient tends to remain in the lipid bilayer that is in the hydrophobic phase. As log *p* increases, the probability of molecule reaching to its critical binding site increases, therefore, crossing of lipid bilayer increases. Another alignment independent topological descriptor *T_2_O_1* and *T_2_O_6* contributed negatively to the model developed by MLR method and is inversely proportional to the activity. *T_2_O_1* and *T_2_O_6* which contributed negatively represent the count of the number of double-bonded atoms separated from oxygen atom by one bond distance and six bond distances respectively in a molecule. The negative contribution of these two descriptors suggests that lower values results in good inhibitory activities while greater values of *T_2_O_1* and *T_2_O_6* results in reduced inhibitory activities. 2D-QSAR study reveals that the physicochemical descriptor plays a pivotal role and is the key descriptor.

In the model generated by PLS method, Alignment Independent (AI) descriptors *T_C_Cl_4* and *T_T_O_5* and *chiV3Cluster* exhibit positive contribution while *T_N_O_5* and *SsOHcount* shows negative contribution. *chiV3Cluster* signifies atomic valence connectivity index (order 3) [Hall 1991]. Molecules with higher values of *T_C_Cl_4* and *T_T_O_5* and *chiV3Cluster* will show good inhibitory activity and molecules with higher values of *T_N_O_5* and *SsOHcount* will exhibit reduced inhibitory activity. These descriptors (Alignment Independent) can be calculated as discussed in Baumann's paper [27]. This descriptor specifies the total number of oxygen atoms linked with one single bond. The descriptor *SdOcount* denotes the total number of hydroxyl group linked with one single bond.

The most significant 2D models of methaniminium derivatives using MLR and PLS method was compared and it was concluded that MLR method generated the better model. The multiple linear regression (MLR) analysis calculated squared correlation coefficient r^2 value 0.8002, predictive squared correlation coefficient q^2 value 0.6107, F test value 13.6181, root mean square error r^2_{se} value 0.1172, predictive squared correlation coefficient standard error q^2_{se} value 0.1636, $pred_r^2$ value 0.5248 and root mean square error predicted ($pred_r^2_{se}$) value 0.1744.

3D-QSAR modeling

To perform 3D QSAR, the dataset of 30 molecules were taken and it was divided into training (22 compounds) and test set (8 compounds). In this, *K_i* values act as dependent variable while all calculated 3D descriptors act as independent variables.

Molecular modeling and alignment of molecules

Conformational search method (grid search) was performed which gives all potential conformations, by changing consistently each

torsion angle of the molecule by increasing it gradually but at the same time maintaining the bond length and bond angles. The conformers with the lowest energy (therefore more stable) were chosen. Template-based alignment method was applied to align all the molecules of the series. For this purpose, the most active compound with minimum energy conformation was used as the template and all the compounds were aligned on that [28].

Computation of field descriptors

In order to perform 3D QSAR, steric and electrostatic field descriptors were computed having cutoffs of 30.0 kcal/mol and 10.0 kcal/mol respectively. The Gasteiger-Marsili [29] charge type was chosen. Distance-dependent dielectric function was selected; keeping the value of dielectric constant 1.0. A carbon atom with charge 1.0 was set as the probe. Using this, electrostatic (1040 descriptors), steric (1040 descriptors) and hydrophobic descriptors (1040 descriptors) were calculated and altogether 3120 descriptors were obtained. From this set of descriptors, invariable columns were removed as they have no contribution towards QSAR.

K-nearest neighbour molecular field analysis (kNN-MFA)

The foremost need of kNN-MFA is an alignment of molecules. After that, a rectangular grid is generated all around the molecules. Then at the lattice points of the grid, electrostatic and steric energies are calculated which are further employed for relationship generation with the help of kNN method to determine distances between molecules [30].

k-NN-MFA with stepwise forward-backward (SW-FB) variable selection method

Forward and backward stepwise selection method was employed to generate k-NN-MFA models with cross-correlation limit fixed to 1.0 and q^2 was taken as term selection criteria. The values which were set for F-test "in" was 4.0 and F-test "out" as 3.99. Variance cut off was kept at 0.0 kcal/mol \AA^2 and scaling was set as Auto Scaling. For k-Nearest Neighbour parameter setting, the number of maximum neighbors were set to 5, numbers of minimum neighbors were set to 2 and distance based weighted average was chosen as prediction method.

3D QSAR model generation and interpretation

K-NN-MFA method was employed for the generation of 3D-QSAR models. For this purpose, the dataset of 30 compounds was used which was further divided into training and test set. The k-NN-MFA models (3-4) were obtained utilizing a training set of 22 compounds and test set of 8 compounds. The hydrophobicity (H), electrostatic (E) and steric (S) descriptors particularize the regions, whereby changing the structure of the compound of the training set; the activities of the compound may increase or decrease. The descriptors are associated with numbers which correspond to its place in the 3D MFA grid. Many statistically significant models were generated using stepwise forward, backward variable selection method, from which model 4 meets the selection criterion at its best. The criterion of selection of the best model was internal and external predictive ability of the model. This was achieved by q^2 which represents the internal predictive ability of the model and by $pred_r^2$, which anticipate the activity of an external test set.

Table 6-i: It shows 3D molecular descriptors of training set used in the k-NN-MFA (model 4)

Compound	E_513	S_434	E_1897	H_1619
10	-0.05751	-0.00737	0.261705	0.362097
16	-0.07086	-0.02495	0.214728	0.420844
18	0.052579	-0.04344	0.176186	0.345001
23	-0.05253	-0.08391	0.187079	0.288696
26	-0.11056	30	0.116523	0.421608
27	0.056585	-0.03845	0.075065	0.386805
32	-0.00559	-0.01633	0.195233	0.402994
40	-0.4109	-0.02305	0.188377	0.358622
41	0.038644	-0.11408	0.068592	0.327683
42	0.210907	-0.02251	0.172865	0.383975
43	-0.00752	-0.03611	0.142588	0.369393
44	-0.15747	-0.04573	0.082705	0.342664

45	0.143883	-0.04939	0.140757	0.362183
47	0.367139	-0.13843	-0.14734	0.313056
48	0.025765	-0.10398	-0.06093	0.322267
50	-0.00545	-0.012	0.082555	0.384552
51	0.182347	-0.02065	0.129918	0.316739
52	-0.12303	-0.0318	0.081317	0.349608
53	0.016184	-0.0914	0.079591	0.371672
55	-0.04311	-0.11282	0.095685	0.359617
6	-0.06325	-0.02067	0.1859	0.313061
8	0.066054	-0.02051	0.124146	0.397901

Table 6-ii: It represents 3D molecular descriptors of test set used in the k-NN-MFA (model 4)

Compound	E_513	S_434	E_1897	H_1619
13	0.027768	-0.0115	0.099411	0.414873
15	0.045428	-0.04021	0.19768	0.526284
21	-0.16166	-0.04177	0.286382	0.41448
33	0.191621	-0.02135	0.159468	0.351299
36	0.359089	-0.06424	0.027854	0.323807
39	0.166087	-0.09394	0.161573	0.369388
46	0.170558	-0.02543	0.07286	0.358882
49	0.026955	-0.01761	0.085779	0.375608

Analysis of statistical parameters suggests that the maximum of test set compound of Amino (3-((3, 5-difluoro-4-methyl-6-phenoxy pyridine-2-yl) oxy) phenyl) methaniminium derivative is less than the maximum of the training set of compounds. The minimum binding affinity of training set compounds is also less

than the minimum of test set compounds. The average of test set compounds is higher than the average of train set compounds which denotes that comparatively more active compounds are present in test set than in train set. The statistical parameters are shown in table 8.

Table 7: It shows statistical parameters (uni-column statistics) for biological activity distribution in training and test sets of 3D QSAR (model 4 SW-FB)

Parameters	Training set	Test set
Max.	5.8900	5.7700
Min.	5.2300	5.300
Std. Dev.	0.2283	0.2087
Sum	120.24	43.9200
Average	5.465	5.4900

Models using k-NN-MFA method

Model-3 (SW-FB)

Ki =E_513 (-0.0709,-0.0525) E_2066 (0.0755, 0.0914) E_1359 (-10.0000,-10.0000)

Model-4 (SW-FB)

Ki =E_513 (-0.0575,-0.0525) S_434 (-0.0839,-0.0074) E_1897 (0.1871, 0.2617) H_1619 (0.2887, 0.3621)

The model 4, which is obtained by SW-FB selection method, is chosen as the best model based on an internal prediction. The leave one out cross-validation squared correlation coefficient, q^2 was calculated as 0.8790 which indicated good prediction. It also shows better prognostic power for the external test set having predictive squared correlation coefficient value (predicted r^2) equal to 0.9340

which is equal to 93 % predictive power. Table 3.10 depicts actual and predicted values of test set and training set data by model 4 with their residual values.

The best 3D-QSAR model (model 4) established that electrostatic, steric and hydrophobic interactions contribute majorly in prediction. In this, E_513 and E_1897 are electrostatic descriptors while S_434 is steric descriptor. H_1619 is a hydrophobic descriptor.

The negative value of the electrostatic field descriptor shows that negative electronic potential is needed to enhance the activity, therefore more electronegative groups are favoured in that position. Likewise, negative values of steric descriptors suggest that negative values of steric potential are preferable for activity and less bulkier groups as substitutes are favoured in that particular area. Steric descriptor with positive range denotes that more bulky substituents are favoured in that area.

Table 8: It exhibits summary of statistical parameters for 3D-QSAR models (model 3 and 4) of Amino (3-((3, 5-difluoro-4-methyl-6-phenoxy pyridine-2-yl) oxy) phenyl) methaniminium derivative using similar test set and training set

Statistical parameter	SW-FB selection model 3	SW-FB selection model 4
k Nearest Neighbour	2	2
n	22	22
Degree of freedom	18	17
q^2	0.8475	0.8790
q^2_{se}	0.0886	0.0794
Pred r^2	0.4965	0.9340
pred r^2_{se}	0.1525	0.0540

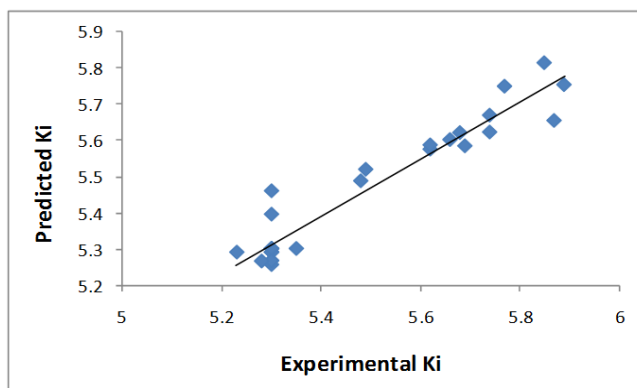


Fig. 5: It shows the graph of experimental versus predicted Ki using model-4 (3D-QSAR)

Table 9: It represents residuals of experimental and predicted inhibitory activities of 3D QSAR for models 3 and 4 containing a test set and training set data

Compound No.	-Log (Ki) for factor Xa experimental	Model 3		Model 4	
		predicted Ki	Residual	Predicted Ki	Residual
10	5.85	5.81448	0.03552	5.81451	0.03549
13	5.77	5.58628	0.18372	5.74955	0.02045
15	5.74	5.59046	0.14954	5.62259	0.11741
16	5.74	5.77543	-0.03543	5.66933	0.07067
18	5.69	5.68094	0.00906	5.58423	0.10577
21	5.68	5.79688	-0.11688	5.62101	0.05899
23	5.66	5.81551	-0.15551	5.60157	0.05843
26	5.62	5.59013	0.02987	5.575	0.045
27	5.62	5.57641	0.04359	5.58694	0.03306
32	5.49	5.47969	0.01031	5.51941	-0.02941
33	5.48	5.48031	-0.00031	5.48747	-0.00747
36	5.35	5.3	0.05	5.3	0.05
39	5.3	5.48592	-0.18592	5.3	0
40	5.3	5.45945	-0.15945	5.3	0
41	5.3	5.44814	-0.14814	5.25546	0.04454
42	5.3	5.49065	-0.19065	5.3	0
43	5.3	5.26515	0.03485	5.39504	-0.09504
44	5.3	5.3284	-0.0284	5.26669	0.03331
45	5.3	5.3	0	5.3	0
46	5.3	5.3	0	5.29025	0.00975
47	5.3	5.3	0	5.3	0
48	5.3	5.31608	-0.01608	5.2901	0.0099
49	5.3	5.28986	0.01014	5.28978	0.01022
50	5.3	5.28997	0.01003	5.45989	-0.15989
51	5.3	5.3	0	5.3	0
52	5.3	5.43031	-0.13031	5.26635	0.03365
53	5.28	5.3	-0.02	5.26519	0.01481
55	5.23	5.3	-0.07	5.28987	-0.05987
6	5.89	5.69988	0.19012	5.75392	0.13608
8	5.87	5.65471	0.21529	5.65457	0.21543

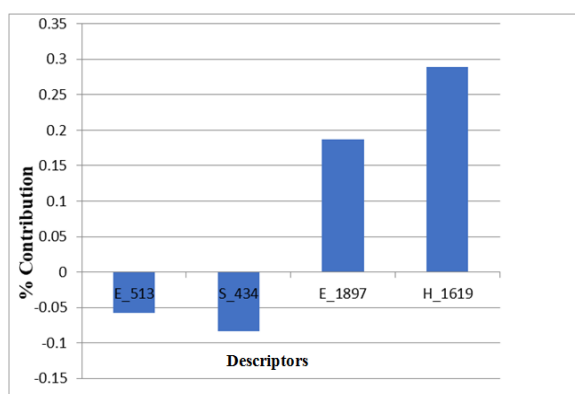


Fig. 6: It shows a contribution chart of the descriptors of 3D-QSAR model (model 4)

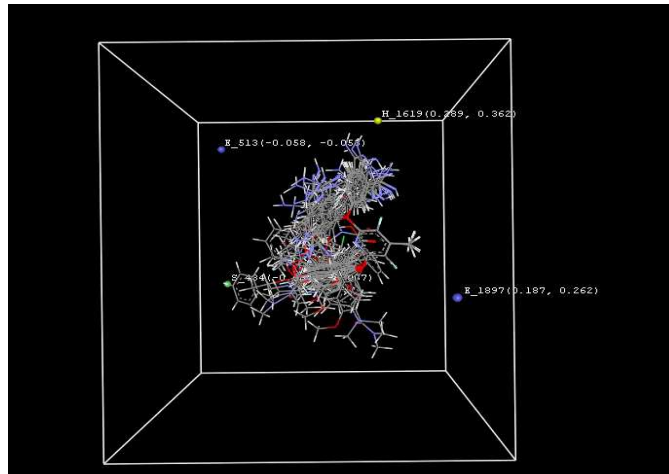


Fig. 7-i: It represents contour plots of 3D-QSAR (model 4) with important hydrophobic, electrostatic and steric fields

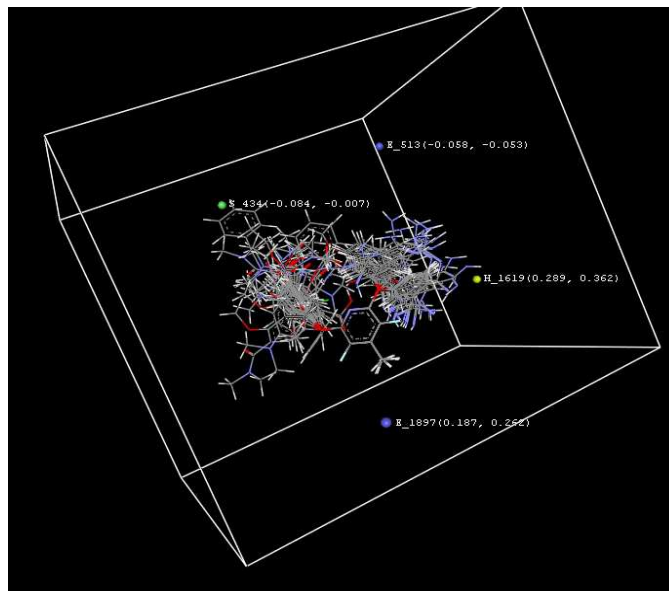


Fig. 7-ii: It represents contour plots of 3D-QSAR (model 4) with important hydrophobic, electrostatic and steric fields

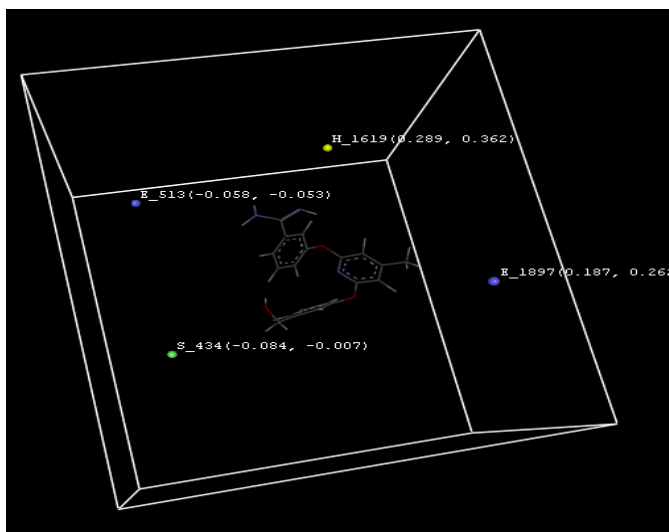


Fig. 8: It shows 3D structure of most active molecule 6

CONCLUSION

On the basis of the best model of 3D QSAR (model 4), the following outcomes were observed for designing of new molecules with regard to electrostatic, steric and hydrophobic fields. The positive value of electrostatic field E₁₈₉₇ (0.1871, 0.2617) denotes that positive electrostatic potential is preferable to enhance the inhibitory activity of compound and therefore substituents with lesser electronegativity values are favourable at that particular place while negative value of electrostatic field E₅₁₃ (-0.0575,-0.0525) shows that for enhancing the activity of compound, the substituent groups with higher electronegativity values are found to be suitable for that particular region. The steric field S₄₃₄ (-0.0839,-0.0074) contributed negatively, which suggests that less bulky substituent groups are preferable in that domain. The positive hydrophobic field descriptors H₁₆₁₉ (0.2887, 0.3621) indicates that hydrophilic groups in that region can raise the activity of the compounds.

Thus, contour plots of KNN-MFA method furnish additional information on the relationship between the structure of compounds and their inhibitory activities which can be employed to construct newer factor Xa inhibitors.

AUTHORS CONTRIBUTIONS

Dr. Smita Suhane is responsible for QSAR analysis and drafting the manuscript.

Dr. A. G. Nerkar has provided advice and guidance during the work.

Dr. Kumud Modi has provided guidance.

Dr. Sanjay D. Sawant has provided guidance during the work.

CONFLICTS OF INTERESTS

There are no conflicts of interest by any of the authors

REFERENCES

- Johnson M, Maggiora GM. Concepts and applications of molecular similarity. Wiley; 1990.
- Hansch, Corwin. Exploring QSAR. Washington DC. American Chemical Society; 1995.
- Hansch, Corwin. The expanding role of quantitative structure-activity relationships (QSAR) in toxicology. *Toxicol Lett* 1995;79:45-53.
- Hansch, Corwin. Antitumor 1-(X-aryl)-3, 3-dialkyltriazenes. 2. on the role of correlation analysis in decision making in drug modification. Toxicity quantitative structure-activity relationships of 1-(X-phenyl)-3, 3-dialkyltriazenes in mice. *J Med Chem* 1978;21:574-7.
- Bradbury SP. Predicting modes of toxic action from chemical structure: an overview. SAR and QSAR in Environ Res 1994;2:89-104.
- Russom CL, SP Bradbury, AR Carlson. Use of knowledge bases and QSARs to estimate the relative ecological risk of agrichemicals: a problem formulation exercise. SAR and QSAR in Environ Res 1995;4:83-95.
- Schultz, Terry W, JR Seward. Health-effects related structure-toxicity relationships: a paradigm for the first decade of the new millennium. *Sci Total Environ* 2000;249:73-84.
- Shi Leming M. QSAR models using a large diverse set of estrogens. *J Chem Inf Comput Sci* 2001;41:186-95.
- Tong W. An integrated computational approach for prioritizing potential estrogenic endocrine disruptors. Proceedings of the International Symposium on Environmental Endocrine Disruptors; 1999.
- Wen, Yuan Hua, Jacob Kalf, Robert Henry Peters. Pharmacokinetic modeling in toxicology: a critical perspective. *Environ Rev* 1999;7:1-18.
- Furie B, Furie BC. Mechanisms of thrombus formation. *New England J Med* 2008;359:938-49.
- Handin RL, Kasper DL, Braunwald E, Fauci AS. Harrison's principles of internal medicine. 16th ed. New York, NY. McGraw-Hill Medical Publishing Division; 2005. p. 337-43.
- Waldo AL. Anticoagulation: stroke prevention in patients with atrial fibrillation. *Med Clin North Am* 2008;92:143-59.
- Fuster V, Moreno PR, Fayad ZA. Atherothrombosis and high-risk plaque: part I: evolving concepts. *J Am Coll Cardiol* 2005;46:937-54.
- Tapson VF. Acute pulmonary embolism. *N Engl J Med* 2008;1037-52. <http://dx.doi.org/10.1136/heart.85.2.229>
- Colman RW, VJ Marder, AW Clowes. Overview of coagulation, fibrinolysis, and their regulation. Hemostasis and Thrombosis: Basic Principles and Clinical Practice Philadelphia; 2006. p. 17-20.
- Hyers TM. Management of venous thromboembolism: past, present, and future. *Arch Intern Med* 2003;163:759-68.
- Hirsh J, O'donnell M, Weitz JI. New anticoagulants Blood; 2005. p. 105-453.
- Weitz JI, Hirsh J, Samama MM. New antithrombotic drugs: American college of chest physicians evidence-based clinical practice guidelines. 8th ed. *Chest J* 2008;133:234S-56S.
- Turpie AG. New oral anticoagulants in atrial fibrillation. *Eur Heart J* 2008;29:155-65.
- Leung LL, Mannucci PM, Landaw SA. Anticoagulants other than heparin and warfarin. *Pharmas Guide Hematology*; 2014.
- Becker RC. Next-generation antithrombin therapies. *J Invasive Cardiol* 2009;21:179-85.
- Franchini M, Mannucci PM. A new era for anticoagulants. *Eur J Intern Med* 2009;365:562-68.
- Böhm M, Sturzebecher J, Klebe G. Three-dimensional quantitative structure-activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa. *J Med Chem* 1999;42:458-77.
- Phillips, Gary. Design, synthesis, and activity of a novel series of factor Xa inhibitors: optimization of arylamidine groups 1, 2. *J Med Chem* 2002;45:2484-93.
- VLife MDS. version 4.3, VLife Sciences Technologies Pvt. Ltd. Pune India; 2008.
- K Baumann. An alignment-independent versatile structure descriptor for QSAR and QSPR based on the distribution of molecular features. *J Chem Inf Comput Sci* 2002;42:26-35.
- Balajee R, Dhanarajan MS. 3D QSAR studies of identified compounds as potential inhibitors for anti-hyperglycemic targets. *Asian J Pharm Clin Res* 2015;8:362-4.
- Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity-a rapid access to atomic charges. *Tetrahedron* 1980;36:3219-28.
- Shen Min. Quantitative structure-activity relationship analysis of functionalized amino acid anticonvulsant agents using k nearest neighbour and simulated annealing PLS methods. *J Med Chem* 2002;45:2811-23.