Marmara Pharmaceutical Journal 21/3: 620-630, 2017
DOI: 10.12991/marupj.323292

RESEARCH PAPER

Investigation of fragment based quantitative regression on a series of substituted chromen-2-one derivatives as FXa inhibitors

Santosh Sahadeo Kumbhar, Prafulla Balkrishna Choudhari, Manish Sudesh Bhatia

ABSTRACT

Factor Xa (FXa), a trypsin-like serine protease, is well-established target for the development of the anticoagulants. Number of molecules were reported as Factor Xa inhibitors but most of them have pharmacokinetic issues. In this present communication, we report development and validation of the group based quantitative structure activity relationship (GQSAR) studies on 48 chromen-2-one derivatives as effective inhibitors of FXa. All the molecules were fragmented into eleven functional fragments (R1, R2, R3, R4, R5, R6, R7, R8, R9, R10 and R11). All the developed GQSAR models were generated using multiple linear regression analysis (MLR). The generated GQSAR models were selected on the basis of statistical data that models having r^2 should be above 0.6 were used to check the external predictivity while the significance of the model was decided on the basis of F value. Developed GQSAR models reveled presence of lipophilic groups on fragment R6 will diminish the bio-activity while at R2 it will lead to increase in bioactivity of molecules. Additionally, minimum number of rotatable bonds at fragments R1 was fruitful for better FXa inhibition activity. The results of GQSAR models may lead to better understanding of design and development of novel FXa inhibitors.

Keywords: FXa, Vlife MDS, Anticoagulant, Quantitative Structure Activity Relationship, GQSAR, Slogp

1. INTRODUCTION

Anticoagulants are used as preventive agents in deep vein thrombosis to prevent blood clots and as prophylactic in patients having mechanical heart valves for lifelong therapy (1). The coagulation cascade is a complex system linking various biomolecular proteins which are known as clotting factors. In recent years these clotting factors are emerged as potential targets for the anticoagulant molecules (2). Current anticoagulant therapy has various side effects like enhanced chance of bleeding, variable patient responses, heparin-induced thrombocytopenia (HIT) and the inability to inhibit clot-bound thrombin are associated with current anticoagulant (3-5). To overcome these limitations, newer direct inhibitors of thrombin and factor Xa are considered to be a better option to indirect clotting factor inhibitors. An important advantage of direct inhibition is that both circulating and clot-bound thrombin can be inhibited (6). FXa is an important target due to involvement at the downstream in the coagulation cascade, as FXa is present at the junction of both extrinsic and intrinsic pathways. Traditionally, prophylactic treatment has been based on

Santosh Sahadeo Kumbhar, Prafulla Balkrishna Choudhari, Manish Sudesh Bhatia
Computational Chemistry Research Lab., Department of Pharmaceutical Chemistry, Bharati Vidyapeeth College of Pharmacy, Kolhapur (MS) India 416013.

Corresponding Author:
Santosh Sahadeo Kumbhar
e-mail: santosh.kumbhar0307@gmail.com, santosh.kumbhar0307@yahoo.in

Submitted / Gönderilme: 31.01.2017 **Revised / Düzeltme:** 22.02.2017
Accepted / Kabul: 23.02.2017

vitamin K antagonists (VKAs), Warfarin, a coumarin derivative has been the drug of choice for the prevention and treatment of arterial and venous thrombotic disorders for more than 60 years (7, 8). As it has narrow therapeutic window and various side effects dose monitoring is required for warfarin. New oral direct FXa inhibitors like rivaroxaban and apixaban anticoagulants are better alternatives for warfarin. Rivaroxaban and apixaban are approved for treatment of venous thromboembolism (VTE) also stroke prevention and systemic embolization in non-valvular atrial fibrillation (AF) (9, 10). The occurrence of patients with AF is growing significantly worldwide, and around 1%–2% of the world population is affected with AF (11-14). QSAR methods have a great significance in modern medicinal chemistry due to ability to directly correlate molecular structure with biological activity of molecules (15, 16). Group based QSAR (GQSAR) is a new approach to scrutinize structure activity relationship based on molecular fragments of congeneric as well as non-congeneric set of molecules (17-20). As the name indicates GQSAR allows developing quantitative relationship between biological activity and various physicochemical descriptors calculated for molecular fragments. Thus the fragmentation of the molecules forms an essential step in order to carry out GQSAR. In this study, we report GQSAR method of QSAR to identify important molecular sites and their properties to aid in the development of novel FXa inhibitors. The dataset of chromen-2-one derivatives developed by Bhatia et al. were utilized to develop fragment based GQSAR models. The developed GQSAR models were found to be useful for framework bounding and optimization of lead molecules as potent inhibitors. In addition to that, the developed GQSAR models also give idea about essential fragment based features this will helpful to medicinal chemist to design combinatorial library (19-21).

2. Materials and methods

2.1. Dataset preparation

Dataset of 48 molecules was taken from literature reported by Bhatia et al. (1) Structure and activity (pIC50) as mentioned in the literature are listed in Table no.1. Structures of pyridyl chromen-2-one derivatives were drawn using 2D draw module of Vlife MDS 4.4. The 2D structures were converted to 3D structures and their energy was minimized on Vlife engine platform in batch. Energy minimization was carried out using Merck molecular force field (MMFF) and Gasteiger charges with 0.01 as convergence criteria (RMS gradient) and analytical gradient type was used (20, 22). Template file was created by replacing atoms at attachment points on a molecule by dummy atom (X). The study was performed on Vlife MDS 4.4 supplied by Vlife Sciences, Pvt. Ltd. Pune, India.

2.2. Calculation of descriptors:

All the optimized molecules were open in Vlife workspace and activity of molecules was imported via insert data command from GQSAR module. Total of 535 2D descriptors were calculated using VLife MDS 4.4 software. Various descriptors like physicochemical, structural, polar surface area, Kappa, individual, chi, chiv, chain path count, cluster, element count, topological, Baumann alignment independent topological descriptors for each molecule under study (23). The template used for the fragmentation of set of molecules shown in (Fig. 1). List of descriptors which were utilized in the present study are described with their description in following Table no 1.

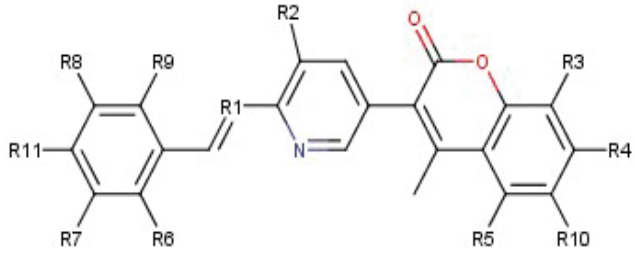


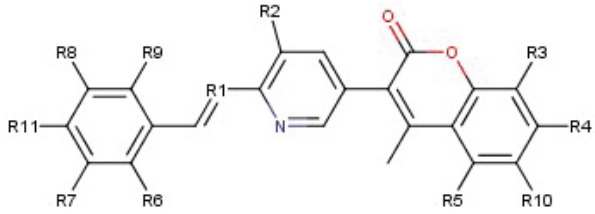
Figure 1. Template used for Fragmentation pattern

2.3. Data selection and building GQSAR model

In the present study dataset of 48 molecules were divided into test and training sets such that 70% of molecules in training set for model building and 30% of molecules in test set for validation of model and there will be uniform distribution of molecules with respect to the bioactivity of molecules shown in Table no. 2. This selection was done by random selection of molecules from dataset. 33 molecules were selected for training set whereas 15 molecules were selected for test set. Multiple linear regression (MLR) method was used together with stepwise (SW) forward-backward variable selection search algorithm was employed for development of the GQSAR model. The GQSAR model with correlation coefficient more than 0.6 were selected and these were used to check external predictivity whereas models showing pred_r^2 below 0.5 were rejected.

Table 1. Descriptors utilized in GQSAR model and their description

Sr. no.	Class of Descriptor	Description
1.	Hydrogens Count	This descriptor signifies number of hydrogen atoms in a compound.
2.	Oxygens Count	This descriptor signifies number of oxygen atoms in a compound.
3.	1Path Count	This descriptor signifies total number of fragments of first order (bonds) in a compound.
4.	Epsilon3	Measure of electronegative atom count including hydrogen atoms with respect to the saturated hydrocarbon (reference alkane) created from the molecule/fragment under consideration.
5.	Mom Inertia Z	This descriptor signifies moment of interia at Z-axix.
6.	Radius of Gyration	This descriptor signifies size descriptor for the distribution of atomic masses in a molecule.
7.	Psi1	A measure of hydrogen–bonding propensity of the molecules and/or polar surface area.
8.	slogp	This descriptor signifies log of the octanol/water partition coefficient.
9.	chiV3Cluster	This descriptor signifies valence molecular connectivity index of 3rd order cluster.
10.	Epsilon4	Measure of electronegative atom count including hydrogen atoms with respect to the saturated hydrocarbon (reference alkane) created from the molecule/fragment under consideration.
11.	SsCH3count	This descriptor defines the total number of -CH_3 group connected with single bond.
12.	XlogP	This descriptor signifies ratio of solute concentration in octanol & water and generally termed as Octanol Water partition Coefficient.
13.	HosoyaIndex	This descriptor signifies the topological index or Z index of a graph is the total number of matching in it plus 1 ("plus 1" accounts for the number of matchings with 0 edges).
14.	SsssNcount	This descriptor defines the total number of nitrogen connected with three single bonds.
15.	MomInertiaY	This descriptor signifies moment of interia at Y-axis.
16.	SaaCHE-index	Electro topological state indices for number of –CH group connected with two aromatic bonds.
17.	Rotatable Bond Count	Number of rotatable bonds.
18.	SaaCHE-index	Electro topological state indices for number of –CH group connected with two aromatic bonds.
19.	Hydrogens Count	This descriptor signifies number of hydrogen atoms in a compound.
20.	Oxygens Count	This descriptor signifies number of oxygen atoms in a compound.
21.	chiV3Cluster	This descriptor signifies valence molecular connectivity index of 3rd order cluster.
22.	1PathCount	This descriptor signifies total number of fragments of first order (bonds) in a compound.



General structure

2.4. Validation of the developed GQSAR model

Validation of GQSAR model is most important and critical step in development of model. Validation of model is done to test the internal stability of model and predictive ability of the GQSAR model. Validation is of two types internal (methods like the methods of least squares fit (r^2), cross-validation (q^2), adjusted r^2 (r^2adj), chi-squared test (χ^2), Root Mean Squared Error (RMSE), bootstrapping and scrambling) and external (using a separate test set of molecules). One of the best method for validation of a model is an external validation method that is evaluation of the QSAR model on a test set of molecules (24).

2.4.1 Internal validation

Internal validation of developed model was carried out by means of Leave-One-Out (q^2, LOO) method (25). To calculate q^2, every molecule in the training set was removed one after the other, and the activity of the removed molecule was predicted from the developed model from the remaining set of molecules. Equation 1 used to calculate q^2 which describes the internal stability of the model.

Eq. 1

$$q^2 = 1 - \frac{\Sigma (y_i - \ddot{y}_i)^2}{\Sigma (y_i - y_{mean})^2}$$

Where, yi and ŷi are the actual and predicted activities of the i^{th} molecule, individually and y mean is an average activity of molecules from training set.

2.4.2 External Validation

For the external validation of the model, activity of individual molecule from the test set was predicted using the model produced from the molecules in the training set. Equation 2 used to calculate the pred_r^2 value as follows,

Eq. 2

$$Pred\_r^2 = 1 - \frac{\Sigma (y_i - \ddot{y}_i)^2}{\Sigma (y_i - y_{mean})^2}$$

Where, yi and *ŷ*i are the actual and predicted activities of the i^{th} molecule, individually and y mean is an average activity of molecules from the training set (25, 26).

3. Results and discussion

3.1. GQSAR studies

In the present study we performed a GQSAR on the coumarin derivatives as anticoagulant agents reported by Bhatia et al.(1). The molecules in the study were randomly divided into the training set and test set. 70% of molecules were selected for training set and remaining 30% of molecules in test set for validation of model. MLR method was used together with SW forward-backward variable selection search algorithm with cross-correlation limit 0.7 was employed for development of the significant GQSAR model. Several GQSAR were models developed and best one were selected based on various statistical parameters as r^2, q^2, pred_r^2, F-test and standard error. The predicted activities of four different GQSAR models A, B, C, and D are shown in Table no. 3.

3.2. Interpretation of QSAR Models

The model was selected on basis of r^2, q^2, pred_r^2, F-test and standard error to evaluate molecular features that govern the anticoagulant potential of the selected derivatives.

GQSAR Model A

The GQSAR models developed based on above statistical parameters model A was more significant having r^2= 0.72, q^2 = 0.64 and pred_r^2 = 0.54. The dataset used for present study and other parameters observed and predicted activity of molecules are given in Table 2. The fitness plot for model A is shown in figure 1, which shows random distribution of predicted versus observed activity of molecules. The model A shows descriptors like R1-SaaCHE-index, R2-slogp, R1-chiV3Cluster, R6-slogp and R1-RotatableBondCount are contributed to activity of molecules.

pIC50= 0.3371+0.0294(±0.0003) R1-SaaCHE-index + 0.0802(±0.0093) R2-slogp+ 0.9600(±0.1277) R1-chiV3Cluster -0.2393(±0.0955) R6-slogp -0.0227(±0.0005) R1-RotatableBondCount

Having n= 33, Degree of freedom = 38, in adding to that, the randomization test shows confidence of >99.9999% (Alpha Rand R^2=0.00001) that indicate the generated model A is not random therefore GQSAR model A is chosen as the best QSAR model.

R1-SaaCHE-index signifies electro-topological state indices for number of –CH group connected with two aromatic

Table 2. Table showing dataset used under study

Sr. No.	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11
1	=N-	H	H	H	H	H	H	H	-OH	H	H
2	=N-	H	H	H	H	H	H	H	H	H	$Cl\text{-}C_6H_5\text{-}$
3	=N-	H	H	H	H	H	H	H	H	H	$-OCH_3$
#4	=N-	H	H	H	H	H	H	H	H	H	H
#5	=N-	H	H	H	H	H	H	H	H	H	$(CH_3)_2N$-
#6	=N-	H	H	H	H	H	$-OCH_3$	$-OCH_3$	H	H	$-OCH_3$
7	=N-	NO_2	H	H	H	H	H	H	-OH	H	H
8	=N-	NO_2	H	H	H	H	H	H	H	H	-Cl
9	=N-	NO_2	H	H	H	H	H	H	H	H	$-OCH_3$
#10	=N-	NO_2	H	H	H	H	H	H	H	H	H
11	=N-	NO_2	H	H	H	H	H	H	H	H	$(CH_3)_2N$-
#12	=N-	NO_2	H	H	H	H	$-OCH_3$	$-OCH_3$	H	H	$-OCH_3$
#13	=N-	H	H	-OH	H	H	H	H	-OH	H	H
14	=N-	H	H	-OH	H	H	H	H	H	H	-Cl
15	=N-	H	H	-OH	H	H	H	H	H	H	$-OCH_3$
16	=N-	H	H	-OH	H	H	H	H	H	H	H
17	=N-	H	H	-OH	H	H	H	H	H	H	$(CH_3)_2N$-
18	=N-	H	H	-OH	H	H	$-OCH_3$	$-OCH_3$	H	H	$-OCH_3$
19	=N-	NO_2	H	-OH	H	H	H	H	-OH	H	H
20	=N-	NO_2	H	-OH	H	H	H	H	H	H	-Cl
#21	=N-	NO_2	H	-OH	H	H	H	H	H	H	$-OCH_3$
#22	=N-	NO_2	H	-OH	H	H	H	H	H	H	H
23	=N-	NO_2	H	-OH	H	H	H	H	H	H	$(CH_3)_2N$-
#24	=N-	NO_2	H	-OH	H	H	$-OCH_3$	$-OCH_3$	H	H	$-OCH_3$
#25	=N-	H	H	H	H	H	H	H	-OH	$-NH_2$	H
26	=N-	H	H	H	H	H	H	H	H	$-NH_2$	-Cl
27	=N-	H	H	H	H	H	H	H	H	$-NH_2$	$-OCH_3$
#28	=N-	H	H	H	H	H	H	H	H	$-NH_2$	H
29	=N-	H	H	H	H	H	H	H	H	$-NH_2$	$(CH_3)_2N$-
30	=N-	H	H	H	H	H	$-OCH_3$	$-OCH_3$	H	$-NH_2$	$-OCH_3$
31	=N-	NO_2	H	H	H	H	H	H	-OH	$-NH_2$	H
32	=N-	NO_2	H	H	H	H	H	H	H	$-NH_2$	-Cl
33	=N-	NO_2	H	H	H	H	H	H	H	$-NH_2$	$-OCH_3$
34	=N-	NO_2	H	H	H	H	H	H	H	$-NH_2$	H
35	=N-	NO_2	H	H	H	H	H	H	H	$-NH_2$	$(CH_3)_2N$-
#36	=N-	NO_2	H	H	H	H	$-OCH_3$	$-OCH_3$	H	$-NH_2$	$-OCH_3$
37	=N-	H	NO_2	H	H	H	H	H	-OH	$-NH_2$	H
38	=N-	H	NO_2	H	H	H	H	H	H	H	-Cl
#39	=N-	H	NO_2	H	H	H	H	H	H	H	$-OCH_3$
40	=N-	H	NO_2	H	H	H	H	H	H	H	H
41	=N-	H	NO_2	H	H	H	H	H	H	H	$(CH_3)_2N$-
42	=N-	H	NO_2	H	H	H	$-OCH_3$	$-OCH_3$	H	H	$-OCH_3$
43	=N-	NO_2	NO_2	H	H	H	H	H	-OH	H	H
44	=N-	NO_2	NO_2	H	H	H	H	H	H	H	-Cl
#45	=N-	NO_2	NO_2	H	H	H	H	H	H	H	$-OCH_3$
46	=N-	NO_2	NO_2	H	H	H	H	H	H	H	H
47	=N-	NO_2	NO_2	H	H	H	H	H	H	H	$(CH_3)_2N$-
#48	=N-	NO_2	NO_2	H	H	H	$-OCH_3$	$-OCH_3$	H	H	$-OCH_3$

#: indicates molecules in test set

bonds which is positively contributing to activity of molecules indicates that –CH at R1 contributes positively in activity. R2-slogp means log of the octanol/water partition coefficient. R2-slogp is directly proportional to bioactivity of molecules which indicates substitutions of lipophilic groups at R2 position of pyridine ring will increase in inhibition of FXa. R1-chiV3Cluster this descriptor signifies that valence molecular connectivity index of 3rd order cluster which also contributing positively in bioactivity of molecules which indicates increase in carbon chain length will potentiate the anticoagulant activity, due to occupation of hydrophobic pockets in the factor XA. R6-slogp means log of the octanol/water partition coefficient. R6-slogp is contributing negatively in bioactivity of molecules which indicates substitution of lipophilic group at R6 position of substituted benzaldehyde will decrease in activity. Lastly descriptor R1-RotatableBondCount means Number of rotatable bonds. R1-RotatableBondCount which is negatively contributing to activity of molecules that means number of rotatable bonds at R1 position should be minimum for better activity of molecules.

Activity distribution plot in which green dots represent for test set whereas red dots represents training set molecules. The y-axis of the plot designate the activity values as in the worksheet shown in fig. 2. Plot indicates that the test set molecule activities place within the range of training set molecules.

Fitness plot for model A is shown in fig. 3. The fitness plot gives knowledge about how well best suited model is and how much it fit to predict the activity of the external molecules in test set. Plot shows that about all molecules are close to the line, which gives assurance about predictive capability of model A.

Contribution Plot for model A shows the comparative contribution of specific descriptors which plays significant role in bio-activity deviation in the developed model shown in fig. 4.

Uni Column Statistics:

Uni-Column Statistics: For Training set

Column Name	Average	Max	Min	StdDev	Sum
pIC50	0.5507	0.8500	0.1600	0.2093	24.2300

Uni-Column Statistics: For Test set

Column Name	Average	Max	Min	StdDev	Sum
pIC50	0.5715	0.7800	0.1800	0.1954	8.4300

Uni-column statistics of model A shows that average of test set is somewhat greater than average of training set which indicates dataset used for current study have comparatively more active molecules than inactive molecules. The value of max of test set less than training set and value of min of test set greater than training set which indicate that test set is derived in between min-max range of training set molecules. Standard deviation of training set is higher than test set this signifies that training set molecules has extensively scattered activity than test set molecules.

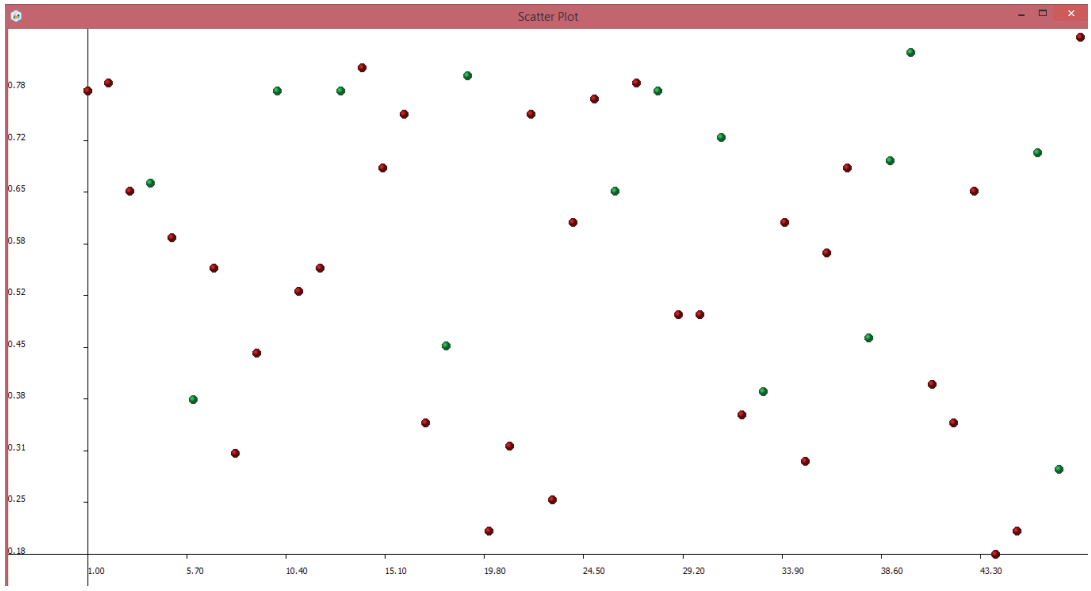


Figure 2. Activity distribution plot of Model A

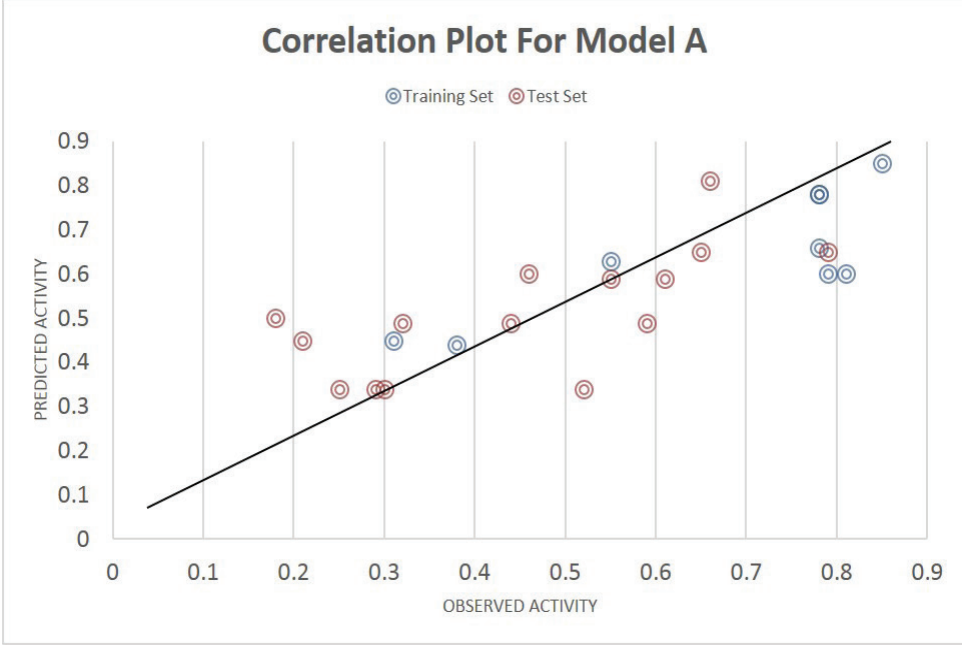


Figure 3. Fitness plot of model A comparison of observed versus predicted activity

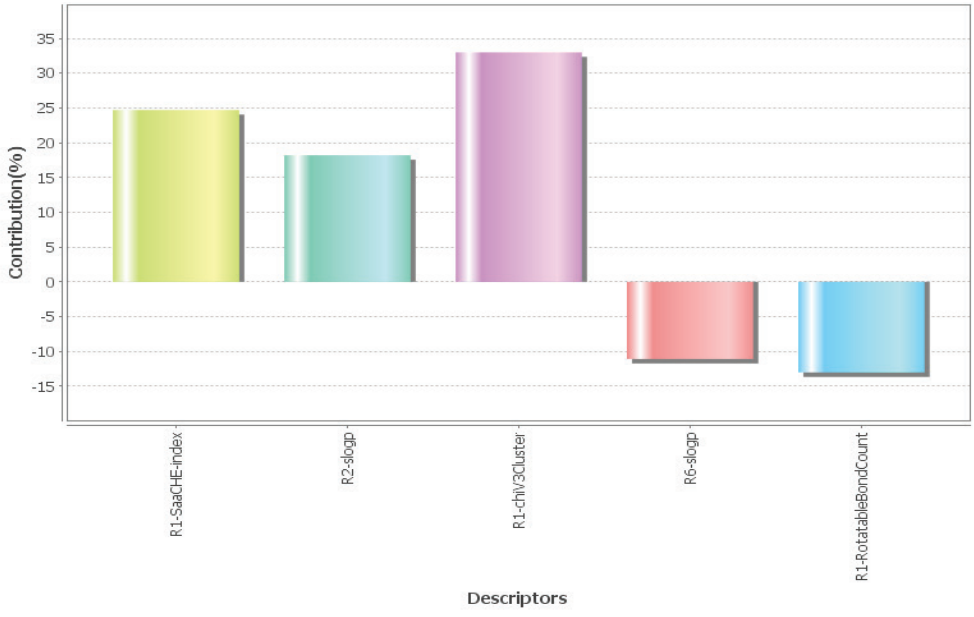


Figure 4. Contribution Plot of developed model A

GQSAR Model B

The obtained GQSAR model B was also statistically more significant with value of r^2 = 0.70, q^2 = 0.61 and pred_r^2 = 0.55. The equation for model B as follows,

pIC50= 0.3384+ 0.1574(±0.0088) R2-HydrogensCount -0.0843(±0.0129) R1-OxygensCount+ 0.5842(±0.1449) R1-chiV3Cluster -0.2565(±0.0996) R6-slogp

With, n= 33, Degree of freedom= 39, Alpha Rand R^2=0.00000 which indicate developed model was good in prediction of activity.

The developed model B shows that descriptor R2-HydrogensCount means number of hydrogen atoms in a compound at R2 site plays vital role and which is directly proportional to regulate activity of molecules. That signifies increase in hydrogen atom count at R2 position will increase

Table 3. Showing statistical parameters of dataset (Observed and Predicted activity)

Compound code	Observed activity pIC50	Predicted activity pIC50			
		Model A	Model B	Model C	Model D
1	0.85	0.85	0.86	0.89	0.85
2	0.78	0.78	0.75	0.83	0.82
3	0.79	0.60	0.60	0.63	0.69
#4	0.65	0.65	0.64	0.68	0.66
#5	0.66	0.81	0.82	0.92	0.83
#6	0.59	0.49	0.49	0.52	0.50
7	0.38	0.44	0.43	0.43	0.40
8	0.55	0.63	0.60	0.64	0.65
9	0.31	0.45	0.44	0.44	0.52
#10	0.44	0.49	0.49	0.49	0.49
11	0.78	0.66	0.66	0.73	0.67
#12	0.52	0.34	0.33	0.33	0.34
#13	0.55	0.59	0.59	0.62	0.57
14	0.78	0.78	0.75	0.83	0.82
15	0.81	0.60	0.60	0.63	0.69
16	0.68	0.64	0.64	0.68	0.66
17	0.75	0.81	0.82	0.92	0.83
18	0.35	0.49	0.49	0.52	0.50
19	0.45	0.44	0.43	0.43	0.40
20	0.8	0.63	0.60	0.64	0.65
#21	0.21	0.45	0.44	0.44	0.52
#22	0.32	0.49	0.49	0.49	0.49
23	0.75	0.66	0.66	0.73	0.67
#24	0.25	0.34	0.33	0.33	0.34
#25	0.61	0.59	0.59	0.62	0.57
26	0.77	0.78	0.75	0.83	0.82
27	0.65	0.60	0.60	0.63	0.68
#28	0.79	0.65	0.64	0.68	0.66
29	0.78	0.81	0.82	0.92	0.83
30	0.49	0.49	0.49	0.52	0.50
31	0.49	0.44	0.43	0.43	0.40
32	0.72	0.63	0.60	0.64	0.65
33	0.36	0.50	0.44	0.44	0.52
34	0.39	0.49	0.49	0.49	0.49
35	0.61	0.66	0.66	0.73	0.67
#36	0.3	0.34	0.33	0.33	0.34
37	0.57	0.59	0.59	0.52	0.46
38	0.68	0.78	0.75	0.73	0.71
#39	0.46	0.60	0.60	0.53	0.57
40	0.69	0.65	0.64	0.57	0.55
41	0.83	0.81	0.82	0.82	0.72
42	0.4	0.49	0.49	0.41	0.39
43	0.35	0.44	0.43	0.33	0.29
44	0.65	0.63	0.60	0.54	0.54
#45	0.18	0.50	0.44	0.34	0.41
46	0.21	0.49	0.49	0.38	0.38
47	0.7	0.66	0.66	0.63	0.56
#48	0.29	0.34	0.33	0.22	0.23

#: indicates molecules in test set

bioactivity of molecules. Another important factor which regulates the activity of molecule is number of oxygen atoms in a compound at R1 site (R1-OxygensCount) which is negatively contributing to activity of molecules. At R1 site of molecule R1-chiV3Cluster which signifies valence molecular connectivity index of 3rd order cluster which is directly proportional to activity of molecules. Last but not least another important factors that governing deviation in the bioactivity of molecules is the log of partition coefficient at R6 substitution site (R6-slogp) which is negatively contributed in bioactivity of molecules. This signifies that substitution of lipophilic group at R6 position of substituted benzaldehyde will contribute to diminution in activity.

GQSAR Model C

The generated model C was also having significant statistical values, $r^2 = 0.68$, $q^2 = 0.43$ and pred_$r^2 = 0.57$. The equation generated for model C as follows,

pIC50= 0.3443+ 0.1897(±0.0117) R2-HydrogensCount -0.1003(±0.0146) R1-OxygensCount+ 0.8291(±0.1753) R1-chiV3Cluster -0.2003(±0.0794) R3-Epsilon4 -0.2565(±0.1168) R6-slogp

With, n= 33, Degree of freedom= 38, Alpha Rand R^2=0.00002 which designate that developed model was good in prediction of activity.

The developed model C shows that the descriptor R2-HydrogensCount signifies that number of hydrogen atoms in a compound at R2 site plays key role and which is positively contributed to regulating the activity of molecules. This means increase in hydrogen atom count at R2 position will increase bioactivity of molecules. At R1 position of molecule descriptor R1-OxygensCount which is inversely proportional to activity of molecules. R1-chiV3Cluster that signifies valence molecular connectivity index of 3rd order cluster which is positively contributed to bioactivity of molecules. Another important descriptor at R3 position which is also inversely proportional in determining bioactivity of molecules is R3-Epsilon4 which signifies Measure of electronegative atom count including hydrogen atoms with respect to the saturated hydrocarbon (reference alkane) created from the molecule/fragment under consideration. That means substitution of electronegative atom at R3 position will diminish the activity of molecules. R6-slogp signifies log of the octanol/water partition coefficient. R6-slogp is contributing negatively in bioactivity of molecules which means that the substitution of lipophilic group at R6 position of substituted benzaldehyde will diminish the activity of molecules.

GQSAR Model D

The GQSAR model D shows the significant statistical values as, $r^2 = 0.68$, $q^2 = 0.59$ and pred_$r^2 = 0.51$. The equation generated for model D as follows,

pIC50= 0.3615+ 0.1634(±0.0173) R2-HydrogensCount -0.0905(±0.0201) R1-OxygensCount -0.0370(±0.0146) R3-1PathCount + 0.3060(±0.0943) R11-Psi1 -0.2588(±0.1162) R6-slogp

With, n= 33, Degree of freedom= 38, Alpha Rand R^2=0.00000 which means developed model was good in prediction of activity.

In the developed model D the descriptor R2-HydrogensCount signifies that number of hydrogen atoms in a compound at position R2 plays an important role in regulating activity and which is positively contributing to bioactivity of molecules. That indicates that increase in hydrogen atom count at R2 position will increase bioactivity of molecules. R1-OxygensCount which is negatively contributing to activity of molecules at R1 position of molecule. Another important descriptor 1PathCount at R3 position also govern the bioactivity of molecule, this descriptor signifies total number of fragments of first order (bonds) in a compound. Which is inversely proportional to activity of molecules means increase in fragments of bonds will reduce the bioactivity of molecules. At R11 site the descriptor R11-Psi1 which is positively governing the activity of molecules. Descriptor Psi1 measures the hydrogen–bonding tendency of the molecules and/or polar surface area. That signifies that polar surface area and hydrogen–bonding tendency of the molecules are directly proportional to bioactivity of molecules. Finally, the presence of descriptor R6-slogp which is inversely proportional to bioactivity of molecules shows the role of substitution of lipophilic group at R6 position in determination of activity of molecules.

4. Conclusion

Present communication is an attempt to identify structural features of chromen-2-one derivatives for this we performed fragment based GQSAR. In the present study GQSAR models were developed on the series of 48 chromen-2-one derivatives having potential FXa inhibition activity. The dataset of 48 molecules were randomly divided in such way that 70%

molecules in training set and remaining 30% molecules were in test set, and GQSAR models were developed with Multiple linear regression (MLR) method coupled with stepwise (SW) forward-backward variable selection search algorithm was employed. Four different models A, B, C and D were developed. Based on statistical parameters r^2, q^2, F-test and standard error models were validated. Model A was found to be good over the other developed models because it fulfils all the parameters and having r^2 value significantly more than other models. The developed model A shows descriptors like R1-SaaCHE-index, R2-slogp, R1-chiV3Cluster, R6-slogp and R1-RotatableBondCount are contributed to activity of molecules. Out of that R6-slogp and R1-RotatableBondCount are negatively contributed in bioactivity of molecules and others are positively contributed to enhance activity. R2-slogp is positively contributed to bioactivity of molecules which indicates substitutions of lipophilic groups at R2 position of pyridine ring will enhances inhibition of FXa. The four developed GQSAR models A, B, C and D are having one common descriptor R6-slogp and which is negatively contributing to activity. Presence of lipophilic group at R6 position of molecule will lessen the activity of molecule. As GQSAR helps in providing site specific clues along with specific descriptor this will helpful in better understanding of drug design. From the designed models it can be concluded that at R6 position if we substitute the lipophilic group the activity of designed molecules will be diminishes. The GQSAR model development is a valuable tool to researchers and medicinal chemist to design and develop potent lead molecule with lesser side effects. Thus, the data obtained from this GQSAR model will helpful for the design and development of novel lead molecules to yield potent FXa inhibitors.

Acknowledgments

The authors are thankful to UGC for funding Major Research Project MRP-MAJOR-PHAR-2013-39160 also thankful to Dr. H. N. More, Principal Bharati Vidyapeeth College of Pharmacy, Kolhapur (M. S.) India for providing facilities to carry out the research work.

Fxa inhibitörü etkili sübstitüe kromen-2-on bileşiklerinin fragment temelli kantitatif regresyon özelliklerinin incelenmesi

ÖZ

Faktör Xa (FXa), tripsin-benzeri bir serin proteaz olup antikoagülan etkili bileşiklerin tasarlanmasında temel alınan önemli bir hedeftir. Literatürde, Faktör Xa inhibitörü olduğu bildirilen birçok bileşiğin farmakokinetik açıdan sorunlu olduğu bildirilmektedir. Bu çalışmada, FXa'yı önemli ölçüde inhibe eden 48 kromen-2-on türevinin grup temelli kantitatif yapı etki ilişkisi (GQSAR) üzerinde çalışılmış ve elde edilen bulgular doğrulanmıştır. Bütün bileşikler onbir fonksiyonel fragmente (R1, R2, R3, R4, R5, R6, R7, R8, R9, R10 ve R11) ayrılarak incelenmiştir. Geliştirilen tüm GQSAR modelleri çoklu lineer regresyon analizi (MLR) kullanılarak oluşturulmuştur. Geliştirilen GQSAR modellerinden r^2 değerleri 0.6'nın üzerinde olanlar istatistiksel yöntemler doğrultusunda seçilmiş ve dış tahmine dayalı analiz yönteminin sağlamasını yapmak amacıyla kullanılmış, oluşturulan modelin anlamlılığı F değeri dikkate alınarak tanımlanmıştır. Geliştirilen GQSAR modelleri; lipofilik grupların fragment R6 üzerindeki varlığının biyoaktiviteyi azalttığını, fragment R2 üzerinde bulunmaları durumunda ise biyoaktivitenin arttığını göstermiştir. Buna ek olarak, fragment R1'de minimum sayıda dönebilen bağ olması durumunun FXa inhibisyonunu arttırdığı tespit edilmiştir. GQSAR modellerinin incelenmesi ile elde edilen sonuçların yeni FXa inhibitörlerinin tasarımı ve geliştirilmesi için yararlı olduğu düşünülmektedir.

Anahtar kelimeler: FXa, Vlife MDS, Antikoagülan, Kantitatif Yapı Etki İlişkisi, GQSAR, Slogp

References:

1. Bhatia MS, Ingale KB, Choudhari PB, Bhatia NM, Sawant RL. Application quantum and physico chemical molecular descriptors utilizing principal components to study mode of anticoagulant activity of pyridyl chromen-2-one derivatives. Bioorg Med Chem 2009; 17: 1654-62.
2. Bohm M, Sturzebecher J, Klebe GJ. 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. Med Chem 1999; 42: 458-77.
3. Hirsh J, Anand SS, Halperin JL, Fuster V. Guide to anticoagulant therapy: Heparin. Circulation 2001; 103: 2994-3018.
4. Hogg PJ, Jackson CM. Fibrin monomer protects thrombin from inactivation by heparin-antithrombin III: Implications for heparin efficacy. P Natl Acad Sci USA 1989; 86: 3619-23.
5. Weitz JI, Hudoba M, Massel D, Maraganore J, Hirsh J. Clot-bound thrombin is protected from inhibition by heparin-antithrombin III but is susceptible to inactivation by

antithrombin III-independent inhibitors. J Clin Invest 1990; 86: 385-91.

6. Walenga JM. An overview of the direct thrombin inhibitor argatroban. Pathophysiol Haemost Thromb 2002; 32: 9-14.
7. Riva N, Lip GY. A new era for anticoagulation in atrial fibrillation. Which anticoagulant should we choose for long term prevention of thromboembolic complications in patients with atrial fibrillation? Pol Arch Med Wewn 2012; 122: 45–53.
8. Hart RG, Pearce LA, Aquilar MI. Meta-analysis: Antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. Ann Intern Med 2007;146: 857–67.
9. Perez A, Eraso L, Merli G. Implications of new anticoagulants in primary practice. Int J Clin Pract 2013; 67: 139–56.
10. Vílchez JA, Gallego P, Lip GY. Safety of new oral anticoagulant drugs: A perspective. Ther Adv Drug Saf 2014, 5: 8-20.
11. Stewart S, Hart CL, Hole DJ, McMurray JJ. Population prevalence, incidence, and predictors of atrial fibrillation in the Renfrew/Paisley study. Heart 2001; 86: 516–21.
12. Colilla S, Crow A, Petkun W, Singer DE, Simon T, Liu X. Estimates of current and future incidence and prevalence of Atrial Fibrillation in the U.S. adult population. Am J Cardiol 2013; 112: 1142–7.
13. Hagens VE, Ranchor AV, Van Sonderen E, Bosker HA, Kamp O, Tijssen JG, Kingma JH, Crijns HJ, Van Gelder IC,RACE Study Group. Effect of rate or rhythm control on quality of life in persistent atrial fibrillation. J Am Coll Cardiol 2004; 43: 241–7.
14. Heeringa J, van der Kuip DA, Hofman A, Kors JA, van Herpen G, Stricker BH, Stijnen T, Lip GY, Witteman JC. Prevalence, incidence and lifetime risk of atrial fibrillation: the Rotterdam study. Eur Heart J 2006; 27: 949–53.
15. Hadjipavlou LD. Review, reevaluation, and new results in quantitative structure-activity studies of anticonvulsants. Med Res Rev 1998; 18: 91-119.
16. Singh Bhadoriya K, Sharma MC, Sharma S, Jain SV, Avchar MA. An approach to design potent anti-Alzheimer's agents by 3D-QSAR studies on fused 5,6-bicyclic heterocycles as γ-secretase modulators using kNN–MFA methodology. Arab J Chem 2014; 7:924-35.
17. Patel HM, Noolvi MN, Sharma P, Jaiswal V, Bansal S, Lohan S, Kumar SS, Abbot V, Dhiman S, Bhardwaj V. Quantitative structure–activity relationship (QSAR) studies as strategic approach in drug discovery. Med Chem Res 2014;12: 4991–5007.
18. Ajmani S, Jadhav K, Kulkarni SA. Group-Based QSAR (G-QSAR): Mitigating interpretation challenges in QSAR. QSAR & Comb Sci 2009; 28:36-51.
19. Ajmani S, Kulkarni SA. Application of GQSAR for scaffold hopping and lead optimization in multitarget inhibitors. Mol Inform 2012;31:473-90.
20. VlifeMDS: Molecular Design Suite 4.4. In., 3.0 edn: Vlife Sciences Technologies Pvt. Ltd., Pune, India. 2015.
21. Choudhari P, Kumbhar S, Phalle S, Choudhari S, Desai S, Khare S, Jadhav S. Application of group-based QSAR on 2-thioxo-4 thiazolidinone for development of potent anti-diabetic compounds. J Mol Struct 2017; 1128: 355-60.
22. Halgren TA. Molecular geometries and vibrational frequencies for MMFF94. J Comput Chem 1996; 17: 553–86.
23. Baumann K. 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.
24. Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK. Validation of QSAR models-strategies and importance. Int J Drug Des Disc 2011; 2: 511-9.
25. Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA) 1: Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 1988; 110: 5959–67.
26. Ajmani S, Agrawal A, Kulkarni SA. A comprehensive structure–activity analysis of protein kinase B-alpha (Akt1) inhibitors. J Mol Graph Model 2010; 28: 683-94.