QSAR Study of Thiophene-Anthranilamides Based Factor Xa Direct Inhibitors

Preetpal S. Sidhu
Department of Chemistry and Biochemistry, University of Wisconsin, Milwaukee, WI 53211, USA
Email: preetpalsidhu@gmail.com

Abstract
QSAR studies were performed to understand the structure activity relationship (SAR) and to build the computational model to predict newer inhibitors with improved potency. In this study, a library of thiophene-anthranilamide based inhibitors of factor Xa was used to develop QSAR model. The library was divided into two sets: Training and Test sets. QSAR Model consists of four descriptors with R-square value of 0.80. Based on the statistical parameters, this model can be used to predict the newer inhibitors with improved pharmacological profile.

Keywords: anti-coagulants, factor Xa, QSAR, Thrombosis

Introduction

Thrombotic diseases such as deep vein thrombosis, stroke and myocardial infarction are the leading cause of death in the western world [1]. Anticoagulants such as Heparin and Coumarin are the only available therapies for treatment and prevention of thrombotic disorders, but these therapies are associated with numerous side effects. Heparin along with Low Molecular Weight Heparin requires parenteral administration and patient monitoring due to increased bleeding risk [2]. Coumarin is an orally available drug but has slower onset of action and narrow therapeutic window. Thus, there is a need for an orally available, safe and efficacious anticoagulant [3]. The most common approach is to develop the direct inhibitors of Factor Xa and thrombin. Factor Xa belongs to the class of trypsin–like serine protease [4]. Factor Xa is located at the convergent point of extrinsic and intrinsic pathways of coagulation. Thus, it plays a major role in coagulation cascade and is one of the major targets to develop anticoagulant. Along with factor Va and calcium, factor Xa forms a prothrombinase complex, which is required to convert prothrombin into active thrombin. Thrombin is the ultimate enzyme in coagulation cascade and has several important procoagulant functions. Therefore, factor Xa inhibitors indirectly prevent the thrombin formation without affecting the existing thrombin. Since factor Xa inhibitors does not affect pre-existing thrombin, it is predicted to cause lesser impairment of hemostasis as compared to direct thrombin inhibitors [5]. Several selective inhibitors of thrombin and factor Xa were discovered. Recently, allosteric inhibitors of thrombin and factor Xa were discovered [6-14], but due to lack of understanding of allosteric nature and shallow binding region of these sites, the designs of these inhibitors are very difficult [6].

A large amount of time and resources were devoted to the discovery of selective inhibitors of factor Xa which are orally bioavailable [15, 16]. The objective of current QSAR study is to uncover some of the structural parameters which govern the factor Xa inhibition activity. Understanding the underlying physics behind interaction of ligand with key structural features of the binding site of a target is necessary for design and development of an inhibitor. QSAR model will not only help in better understanding the structure activity relationship of any class of inhibitors but also helps in designing the more active inhibitors.

To further investigate our drug discovery approach, we designed a computational model which can understand the pharmacological data and relate it to structural and electronic

features of the inhibitors. Beside correlation, this model can predict the new inhibitors with better pharmacological profile. To do so, we used the QSAR approach to make a predictive model. As we know that QSAR technique is most often used to build such predictive model for correlation between the structure and the pharmacological character of a ligand [17]. Large numbers of QSAR techniques have been developed over the past several years. The accumulation of information about 3-dimensional structural features of molecules led to the development of 3D descriptors and associated 3D QSAR models [18]. Some of the examples of these 3D descriptors are 3D shape descriptors used in the molecular shape analysis and 3D steric and electronic descriptors used in comparative molecular field analysis (CoMFA) [19]. One of the common differences between the 3D QSAR and traditional QSAR is the number of descriptors are increasing exponentially and hence the multiple regressions are inadequate. Advanced mathematical tools like principal component analysis [20], partial least square [21], machine learning algorithms22 and many more are adequate tools to deal with the advanced problems. CoMFA uses the combination of power of graphics and PLS technique. It has a large number of applications in the field of medicinal chemistry [23]. The traditional QSAR uses a very small number of parameters as compared to the number of compounds in the data set, whereas CoMFA uses a large number of parameters [24]. One of the advantages of 3D QSAR is that it is not limited by use of compound from a congeneric set and hence, the data set can compose of divergent molecular structures. One of the advantages of traditional QSAR is that it overcomes the underlying assumption of 3D QSAR that the predicted conformation is the bioactive one. The traditional QSAR effectively eliminates the task of generation of a large number of conformations. It usually generates the conformation that is most likely to bind to the receptor of the target. Traditional QSAR can be easily adapted to the database search, and similarity or dissimilarity search. It can be easily automated and lacks the computational complexity.

The QSAR model developed in this study consists of both topological and electronic state descriptors as calculated by commercially available MDL software. MDL has the ability to calculate the large number of descriptors. The total number of descriptors calculated by MDL is 138. Although MDL calculates a large number of descriptors, all descriptors do not provide the positive contribution toward the correlative value of the predictive model. Hence, a procedure is required to select the descriptors that have a meaningful contribution for correlative value. This selection is done on the basis of “F to enter” value. The higher F to enter value, the more contribution it makes for predictive model. The successful accomplishment of any QSAR study is the prediction of a lead molecule. To do so, we have devised the ongoing 2D QSAR based database search program. Based on these studies, new lead molecules will be predicted. The preliminary QSAR study developed the model with good predictive value.

Experimental section

For this study, we selected 37 previously reported inhibitors of factor Xa. The biological activities for these compounds were measured by Berlex Biosciences. The data set includes the derivatives of thiophene-Anthranilamides based inhibitors having substitutions on the side chain. The competitive binding of each inhibitor was reported as Ki,app value. The method for measurement of Ki,app value is reported previously. MDL Routine: All the structures were made using the Sybyl 7.3 and minimized by the energy minimization tool. All the structures were saved as individual mol2 file format. A new database was created and filled with the minimum information required for the software to calculate and make predictive model. This set of information was molecule

name, location of molecule and biological data in order as suggested by the MDL guide. All the descriptors of each inhibitor were calculated and entered into the database. Using the stepwise regression analysis, regression was started. Using genetic algorithm, the weights of different descriptors were calculated and reported as the F to enter value. Higher the F to enter value, the more the correlation to the biological activity. Then the stepwise one descriptor was added, the equation was calculated and the R-square value was provided. These steps were repeated until 4 descriptors were added in the equation.

Result and Discussion

The method employed in this study was described in the Experimental Section. Table 1 lists the experimental vs calculated Kd values. Structural and biological data were adapted from Thiophene_Anthranilamides as Highly Potent and Orally Available Factor Xa Inhibitors (J. Med. Chem. 2007, 50, 2967). All the biological data was performed by the same group using the same assay method. This eliminates the chances of error due to use of different methods and different laboratory setting. The care has been taken to report the correct biological data. The R-square value of 0.8016 was obtained using step-wise regression. The statistical data for the QSAR studies are summarized in Table 2 and Figure 1. The following model was generated using the stepwise regression analysis.

Kd = 0.4657*SdaaN - 1.806*knotp - 9.117*xvch6 + 1.647*SdsCH_acnt - 4.25998

The model uses four descriptors to generate the equation to predict the Kd value for new molecules. The regression quality for the model was calculated by MDL software. The R-square value is 0.8016. The calculated standard error of estimation is 0.3789. The F-statistic value is 36.37, P-value is 3.486E-012 and multiple Q-square value is 0.651. The cross validation RSS value is 9.092. The training set was very well described by the regression equation, which was statistically very significant. Cross-validation shows that the constructed model can be used with care to predict the value of Kd for the new molecules. MDL software has the capability to run the cross validation test upon the same database to test the constructed model. Beside the cross validation, the randomization test was also performed to check the validity of selected descriptors set. 100 independent randomization runs were performed using the variable value. Mean R2K value is 0.1073 and mean square deviation for R2k value is 0.05413. The range of values for R2k is between 0.01282 and 0.264. The four descriptors were found to show strong correlation with the biological activity (Kd value for inhibition of factor Xa) and structural features of the inhibitors. The first descriptor is SdaaN which belongs to the atom type E-state descriptor. Each atom is assigned to a particular type of atom or hydride group. In this case, the nitrogen atom in the aromatic ring forms a double bond. The second descriptor is KnotP, which belongs to the class of simple molecular connectivity chi descriptor. KnotP describes the difference between chi cluster 3 and path/cluster 4. This descriptor defines the degree of skeletal branching and pattern of adjacency. The third descriptor is SdsCH_acnt and fourth descriptor is xvch6. They belong to class of valence molecular connectivity chi descriptors. Molecular connectivity chi descriptors provide a quantitative description of the molecular structure build from a graph of the molecule. From the regression analysis, the weights of different descriptors were found, and based on their weights; the structural fragments that were important for the biological activity are modeled to build the new molecule. This descriptor particularly describes the position and influence of nitrogen in the six membered ring systems.

The results from the QSAR model and comparison with the structure of inhibitors show good correlation between the biological activity and structural features. The KnotP descriptor has highest F to enter value. This

means that the KnotP descriptor has good correlation with biological activity. As we can see from the series of compound 10a to 10p, inhibitors having fragment of chi path-cluster 4 have more potency than chi cluster 3 and straight chain containing compounds. Inhibitors 10o and 10p have the highest potency compared to others. The same is true in the series 12a to 12f. Most of the inhibitors in this series have chi cluster-path 4 fragments and hence, they all were more potent than series 10 inhibitors. The inhibitors 13a, 13b and 13c show good example of agreement with the QSAR model. As the chi order increases, the activity decreases. Most of the inhibitors having chi cluster-path 4 fragment are highly potent. SdaaN descriptor has the second highest F to enter value. This means that inhibitors having nitrogen containing aromatic ring in which nitrogen has one double bond were more potent. The inhibitors in series 13 were a good example of this correlation. Most inhibitors in this series have a fragment belonging to the SdaaN group. Some of the compounds in series 17 also have same E-state fragment. Xvch6 was the third most important descriptor which belongs to the class of valence 6th order chain chi descriptors. With regard to structure, the compounds having a six member ring containing a nitrogen atom have some kind of correlation with the biological activity. In this case, all the inhibitors having this fragment were less potent. These inhibitors show inverse correlation with the biological activity.

Table 1: Experimental vs calculated value K_d along with the residual value for direct factor Xa inhibitors. Compounds numbers are adapted from original research article (*J. Med. Chem.* **2007**, *50,* 2967).

	K_d	K_d(calc)	Residual [plot]		K_d	K_d(calc)	Residual [plot]
10i	-0.34242	0.378769	-0.7212	**12d**	1.30103	1.53446	-0.2334
13c	-0.34242	-0.327239	-0.01518	**12e**	1.4437	1.94775	-0.5041
10g	-0.14613	0.335242	-0.4814	**17d**	1.46852	1.44885	0.01967
10d	-0.14613	0.209056	-0.3552	**13g**	1.50864	1.64473	-0.1361
2	0	-0.122913	0.1229	**13i**	1.58503	1.55553	0.0295
10a	0	0.513109	-0.5131	**12b**	1.58503	1.48281	0.1022
10h	0.283997	0.751382	-0.4674	**13j**	1.61979	1.4463	0.1735
3b	0.366532	0.45434	-0.08781	**12f**	1.65758	1.31855	0.339
10c	0.39794	0.424958	-0.02702	**17j**	1.65758	1.60757	0.05001
10l	0.39794	0.154996	0.2429	**17p**	1.69897	1.4463	0.2527
3a	0.443697	-0.0658281	0.5095	**12c**	1.69897	1.53446	0.1645
10f	0.49485	0.967284	-0.4724	**10p**	1.72125	0.947476	0.7738
10n	0.585027	0.767489	-0.1825	**13b**	1.92082	1.5355	0.3853
10j	0.619789	0.534182	0.08561	**13h**	1.92082	1.15205	0.7688
10e	0.657577	0.686849	-0.02927	**17o**	2.09691	1.9853	0.1116
10b	0.677781	0.424958	0.2528	**13a**	2.1549	1.4463	0.7086
10k	0.744727	0.534182	0.2105	**17l**	2.22185	2.25398	-0.03213
3c	0.79588	0.511425	0.2845	**17m**	2.30103	2.40463	-0.1036
13f	0.920819	1.55553	-0.6347	**17n**	2.30103	2.33105	-0.03002
13d	0.920819	1.26522	-0.3444	**17g**	2.39794	2.39794	-4.44E-16
13k	1.22915	1.4463	-0.2172				

Table 2: Validation Statistics for QSAR model

R-Square	Standard Error	Q-square	RSS
0.80	0.3789	0.651	9.09

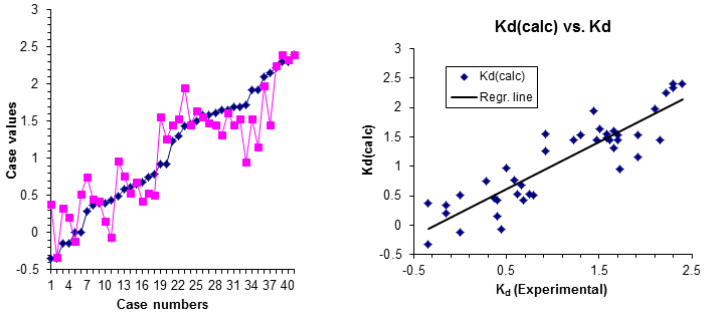

Figure 1. Left – Graphical presentation of experimental vs calculated K_d values. Right - Correlation of calculated vs experimental K_d values.

Table 3: Correlation matrix for computed QSAR Model
Mean and standard deviation

	K_d	knotp	xvch6	SdsCH_acnt	SdaaN
Mean	1.093	-3.143	0.05117	0.02439	0.2246
St.Dev.	0.807	0.2849	0.02009	0.1562	0.6108

Correlation matrix

	K_d	knotp	xvch6	SdsCH_acnt	SdaaN
K_d	1	-0.7117	-0.4285	0.2588	0.4776
knotp	-0.7117	1	0.1896	0.08634	-0.1668
xvch6	-0.4285	0.1896	1	-0.07025	-0.1654
SdsCH_acnt	0.2588	0.08634	-0.07025	1	-0.05887
SdaaN	0.4776	-0.1668	-0.1654	-0.05887	1

Conclusion

The current model developed by us shows good correlation of biological activity with the structural feature of inhibitors. The model is robust and cross validated. Hence, this model can be used of design the new molecules with care. However, there were variations on only one side chain of the molecules. In order to make the model more robust, the inhibitors with variation at different side chain are necessary to be included. Besides using the molecules having the same basic scaffold, a different scaffold must be included but care should be taken on binding site of the molecules.

Acknowledgement

I would like to thank Drs. Desai and Kier for valuable suggestion and Institute of Structural Biology and Drug Discovery, VCU, Richmond, VA for computational resources.

References

(1) Vlasuk, G. P. Direct factor Xa inhibitors. In *New Therapeutic Agents in Thrombosis and thrombolysis*, 1st ed; Sasahara, A. A., Loscalzo, J., Eds.; Marcel Dekker: New York, 1997; pp261.

(2) Kelly, A. G.; Marzee, W. M.; Harker, L. A. Hirudin interruption of heparin-resistant arterial thrombus formation in baboons, *Blood* **1991**, *77*, 1006.

(3) Hirsh, J.; Poller, L. The International Normalized Ratio. A guide to understanding and correcting its problems. *Arch. Int. Med.* **1994**, *154,* 282.

(4) Kaiser, B. Factor Xa-A promising target for drug development. *Cell Mol. Life Sci.* **2002**, *59,* 189.

(5) Samama, M. M.; Gerotiafas, G. T. Evaluation of the pharmacological properties and clinical results of the synthetic pentasaccharide (Fondaparinux). *Thromb. Res.* **2003**, *109,* 1.

(6) Sidhu, P. S.; Allosteric Regulation of Dual Function of Thrombin. *Journal of Pharmacy and Pharmacology* **2013**, *1*, 1-10.

(7) Sidhu, P. S.; Review: One-pot synthesis protocol for highly functionalized benzofuran scaffold. *Journal of Postdoctoral Research* **2014,** *1*, 31-36.

(8) Verghese, J.; Liang, A.; Sidhu, P. S.; Hindle, M.; Zhou, Q.; Desai, U. R.; First steps in the direction of synthetic, allosteric, direct inhibitors of thrombin and factor Xa. *Bioorg. Med. Chem. Lett.* **2009,** *19*, 4126-4129.

(9) Sidhu, P. S.; Liang, A.; Mehta, A. Y.; Abdel Aziz, M. H.; Zhou, Q.; Desai, U. R.; Rational design of potent, small, synthetic allosteric inhibitors of thrombin. *J. Med. Chem.* **2011,** *54*, 5522-5531.