DOI: 10.31276/VJSTE.62(2).24-29

Prediction of inhibition constants of (*R*)-3-amidinophenylalanine inhibitors toward factor Xa by 2D-QSAR model

Thi Bich Van Pham[1], Minh Hao Hoang[2*]

[1]*Department of Chemistry, Faculty of Sciences, Nong Lam University, Ho Chi Minh city*

[2]*Department of Chemical Technology, Faculty of Chemical and Food Technology, Ho Chi Minh city University of Technology and Education*

Received 9 September 2019; accepted 2 December 2019

Abstract:

A coagulation cascade forms through proteolytic reactions and involves different factors. There are two coagulation pathways, including intrinsic and extrinsic mechanisms, which converge by the formation of factor Xa. Factor Xa plays a crucial role in the formation of the complex with factor Va in the presence of calcium ions and phospholipids. This complex converts prothrombin to thrombin, which leads to the formation of a very strong fibrin clot. Much effort has been devoted to the efficient interference of this enzyme cascade by the inhibition of factor Xa due to its important effect. (*R*)-3-amidinophenylalanine inhibitors are known inhibitors of factor Xa reported so far. In the present work, a two-dimensional quantitative structure activity relationship (2D-QSAR) was performed on 50 (*R*)-3-amidinophenylalanine inhibitors (the training set) with respect to their pK_i values toward factor Xa, where pK_i=-logK_i, and K_i is the inhibition constant, to develop a mathematical model that depends on the physicochemical properties of the inhibitors. Partial least squares regression (PLSR) was used to yield a QSAR model containing molecular descriptors that significantly contribute to pK_i values. The statistically significant parameters of the model, such as squared correlation coefficient, R^2=0.834, root mean square error, RMSE=0.210, cross-validated Q^2_{cv}=0.789, and cross-validated $RMSE_{cv}$=0.237, were obtained for the training set. The developed 2D-QSAR model was applied to predict the pK_i values of the 62 inhibitors. Furthermore, the reliability of the model was also confirmed via statistically significant parameters obtained from validation on an external set.

Keywords: **coagulation cascade, descriptors, factor Xa, (*R*)-3-amidinophenylalanine inhibitors, 2D-QSAR.**

Classification number: **2.2**

Introduction

Blood coagulation can be a beneficial response of human body that decreases the amount of bleeding by forming blood clots. These clots play an important role in the sealing of blood vessels to prevent injury from excessive bleeding. However, blood clots can become harmful when they gather together into a compact mass. The presence of large blood clots can cause congestion of blood flow to the body's organs. As a consequence, the supply of oxygen to the organs, especially the brain or heart, is restricted. This leads to a stroke or heart attack. There are two mechanisms leading to coagulation: the contact activation (intrinsic) and tissue factor (extrinsic) pathways [1]. In general, these two pathways occur over several consecutive steps leading to an activation of factor X to factor Xa ("a" activated). Therefore, factor Xa is located at the junction between these two coagulation pathways. In the extrinsic mechanism, factor Xa and factor Va form a complex in the presence of calcium ions and phospholipids. This complex then converts prothrombin to thrombin, which leads to the formation of a very strong fibrin clot [2, 3]. An abnormal clot that forms in a vein may result in pain and swelling, and in many cases, this clot can cause disability and death.

*Corresponding author: Email: haohm@hcmute.edu.vn

Due to the pivotal role of factor Xa to fibrin formation, several great efforts have been made to suppress the coagulation cascade by inhibition of this enzyme. A number of series of novel inhibitors toward factor Xa have been discovered such as mono-benzamidine, non-benzamidine, and diamidino derivatives. These inhibitors have displayed high affinities in *in vitro* and *in vivo* experiments [4]. (*R*)-3-amidinophenylalanine inhibitors were found to represent promising new selective inhibitors of factor Xa due to their hydrophobic interactions with factor Xa [5, 6].

Many drug molecules are enzyme inhibitors and their inhibitory activity is characterised by the inhibition constant, K_i. When an enzyme (E) binds to an inhibitor (I) to form an enzyme-inhibitor complex (EI), E + I $\leftrightarrow$ EI, where K_i is defined as an equilibrium constant such that K_i=[EI]/[E][I], where [E], [I], and [EI] are the equilibrium concentrations of the enzyme, inhibitor, and enzyme-inhibitor complex [7]. A high K_i value ensures that a drug will have high inhibitory activity.

The two-dimensional quantitative structure-activity relationship (2D-QSAR) has seen wide application in the field of medicinal chemistry for many years. This method presents a quantitative relationship between the chemical response (inhibitory activity/toxicity/binding affinity) of a molecule and its physicochemical properties via a mathematical equation [8]. The QSAR method helps to screen new drug candidates, thus avoiding costly trial and error experiments in synthesis and biological screening. In the present attempt, we developed a mathematical model that provided a quantitative relationship of the binding affinity (e.g., pK_i) of (*R*)-3-amidinophenylalanine inhibitors toward factor Xa, a crucial enzyme in the clotting cascade. The quantitative relationship was presented by a mathematically linear equation that depends on molecular physicochemical properties (descriptors) of (*R*)-3-amidinophenylalanine inhibitors. The developed 2D-QSAR model was applied to predict the K_i values of 62 inhibitors.

Methodology

Structures of (*R*)-3-amidinophenylalanine inhibitors and their experimental pK_i=-logK_i values were obtained from the literature [9] (Table 1). Chemical structures were drawn and optimized energy in Molecular Operating Environment (MOE) 2008.10. In order to develop a 2D-QSAR model, a training set including 50 (*R*)-3-amidinophenylalanine inhibitors was randomly chosen in MOE 2008.10. The remaining inhibitors (12 molecules) were used as a testing (external) set. One hundred and eighty-four (184) two-dimensional (2D) descriptors were numerically calculated by MOE software. By using Rapidminer 5.0, the descriptors showing zero value, low correlation with binding affinity (<0.07), and high intercorrelation themselves (>0.9) were removed to select the most significant descriptors for the 2D-QSAR model. In addition, Weka 3.6 software, QuaSAR-Contigency, and Principle Components in MOE 2008.10 were also employed to select the best descriptors to establish the QSAR model. Then, partial least squares regression was used to develop a mathematical equation.

Results

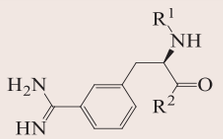
2D-QSAR model

Descriptors are the physicochemical properties of each molecule that characterize its chemical structure and they take on numerical values [8]. After the irrelevant descriptors were omitted, PLSR was employed to develop a mathematical QSAR model that describes a quantitative relationship between the descriptors of (*R*)-3-amidinophenylalanine inhibitors with their pK_i values. The estimated QSAR model is shown below:

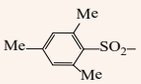
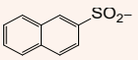
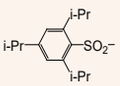
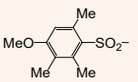
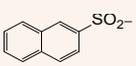
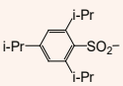
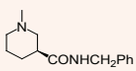
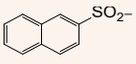
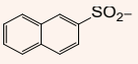
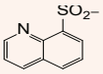
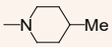
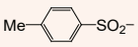
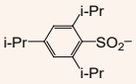
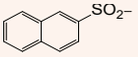
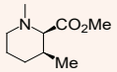
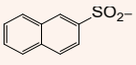
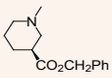
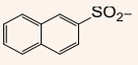
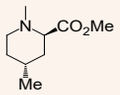
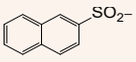
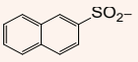
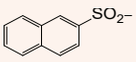
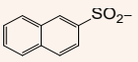
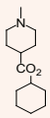
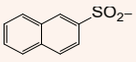
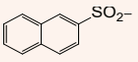
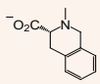
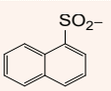
$$pK_i = 3.73958 - 1.14732 \times b\_ar + 0.56128 \times PEOE\_VSA\_POS + 1.16326 \times SlogP\_VSA6 + 2.08858 \times SMR\_VSA5$$

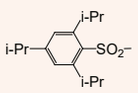
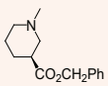
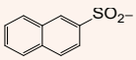
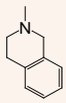
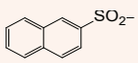
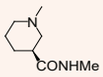
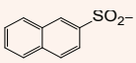
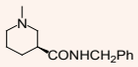
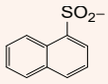
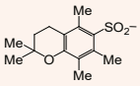
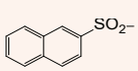
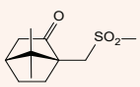
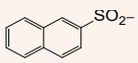
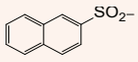
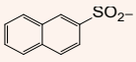
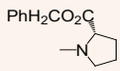
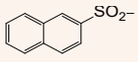
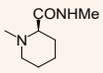
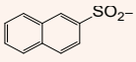
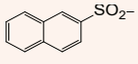
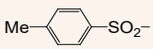
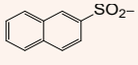
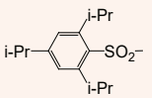
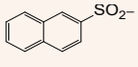
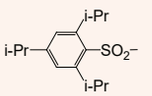
where b_ar is the number of aromatic bonds, PEOE_VSA_POS is the total positive van der Waals surface area, SlogP_VSA is the logarithm of the *n*-octanol/water partition coefficient, and SMR_VSA is the molecular refractivity. The training set was randomly selected from 62 inhibitors to develop the 2D-QSAR model. The model with statistically significant parameters was chosen as the best model. Several training sets were used to develop the 2D-QSAR models. Unfortunately, they gave statistically insignificant R^2, RMSE, Q^2_{cv}, and $RMSE_{cv}$ values. Therefore, those models were not selected for further analysis.

Table 1. Chemical structures of 62 (*R*)-3-amidinophenylalanine inhibitors with respect to their experimental (Exp) pK_i values. The predicted (Pre) pK_i values of 62 inhibitors calculated from the 2D-QSAR equation were also added.



(*R*)-3-amidinophenylalanine inhibitors

N⁰	R¹	R²	Exp pK_i	Pred pK_i	N⁰	R¹	R²	Exp pK_i	Pred pK_i
1		—NHMe	3.194	3.638	32			4.886	4.689
2			5.824	5.494	33		—NHMe	3.721	3.670
3			4.456	4.280	34			5.602	5.909
4			4.337	4.554	35			4.658	4.724
5			3.745	4.243	36			4.119	4.176
6		—NHMe	4.959	4.979	37			4.444	4.461
7			5.066	4.898	38			4.237	4.461
8			3.886	4.270	39			4.319	4.272
9			4.42	4.422	40			5.119	5.126
10			4.367	4.467	41			4.382	4.454
11			4.77	4.386	42			4.886	5.345
12			4.523	4.272	43			4.387	4.196
13			4.796	4.606	44			3.921	3.969
14			4.119	4.484	45			5.114	5.048

15			6.046	5.908	46			4.745	4.453
16			4.481	4.484	47			4.77	4.899
17			4.377	4.293	48			4.097	4.161
18			4.363	4.335	49			4.854	4.560
19			4.357	4.335	50			3.638	3.869
20			4.569	4.486	51*			4.699	4.833
21			4.244	4.571	52*			4.658	4.992
22			4.62	4.428	53*			4.398	3.877
23			4.569	4.616	54*			5.699	5.338
24			4.42	4.421	55*			5.585	5.476
25			4.745	4.467	56*			4.000	4.029
26			3.959	4.402	57*			4.721	4.363
27			4.585	4.572	58*			4.194	4.070
28			4.268	4.315	59*			5.131	5.278
29			4.125	4.101	60*			4.387	4.328
30			4.114	4.272	61*			5.092	4.898
31			4.076	4.135	62*			4.284	4.036

*Testing set.

Statistical parameters

The statistical parameters, such as R^2 and RMSE, are important parameters for the selection of the best 2D-QSAR model. A model was chosen with the greatest R^2 (>0.5), while RMSE value must be below 0.5 [8, 10]. The significant values of R^2, RMSE, Q^2_{cv}, and $RMSE_{cv}$ reflect the reliability of the QSAR model. The obtained values are shown in Table 2.

Table 2. The statistical parameters of the established 2D-QSAR model.

	Training set	Cross-validation	Testing set	Total set
N^0	50	50	12	62
R^2	0.834		0.934	0.814
Q^2_{cv}		0.789		
RMSE	0.210	0.237	0.132	0.227

Experimental pK_i vs. predicted pK_i

The pK_i values of the 62 inhibitors were predicted by using the established 2D-QSAR model. The relationship between the experimental pK_i and predicted pK_i is presented in Fig. 1. The fitting equation is given in the top of the Fig. 1.

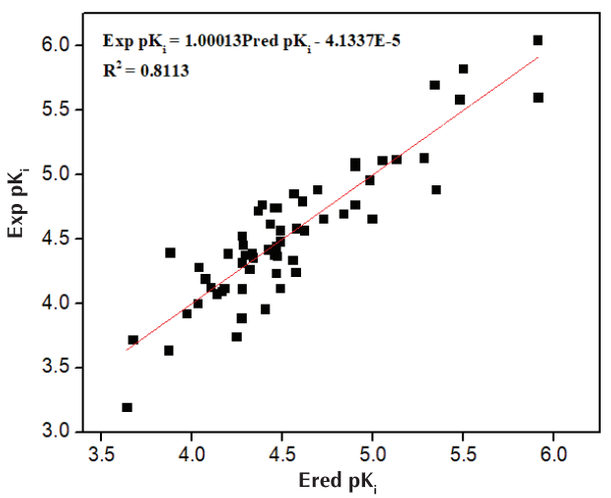


Fig. 1. The plot shows the relationship between the experimental pK_i and predicted pK_i.

Discussion

Subdata set selection to establish a 2D-QSAR model

A data set of 88 inhibitors with respect to their experimental pK_i values was firstly used to perform a QSAR study. As presented in the methodology section, the training set (80%) was selected through their assigned random values (sorted in descending order). After removing the irrelevant descriptors from 184 2D-descriptors, the QSAR model was built. If the first model possessed statistically significant parameters of R^2 and RMSE, the model was used for cross-validations, and if the parameters of Q^2_{cv}, $RMSE_{cv}$ were acceptable, the model was applied to the total set and testing set. Note that the values of R^2, RMSE, and Q^2_{cv}, are dependent on which of the compounds were used in the training set. Therefore, if these values were not statistically satisfied, the first model would not be used further. Consequently, random values of each molecule in 88 inhibitors would be re-calculated and sorted again to select a new training set (the calculation was randomly done by MOE). Then, a second attempt at 2D-QSAR modelling was established based on the new descriptors. This procedure was repeated 8-10 times for the first dataset containing 88 inhibitors and if the validation parameters of R^2, RMSE, and Q^2_{cv} were not statistically significant, the other solutions were taken into account. The reasons behind the statistically insignificant values are the irrelevant selection of descriptors and/or the interfering compounds. Thus, the interfering molecules should be considered. The Z-score values were obtained after cross-validation and the compound outliers to the fit were omitted. As a result, the number of inhibitors will be less. As mentioned earlier, the selection of training and testing sets was random and tthe same procedure was performed to get the statistical parameters. If the results were unacceptable, more interfering molecules were screened via the Z-score until the statistically desired values were obtained from a certain subset of the data. Fortunately, data consisting of 62 inhibitors gave statistically significant parameters for validation.

Molecular descriptors

According to the established equation, pK_i depends on four 2D-descriptors consisting of the number of aromatic bonds, the total positive van der Waals surface area, the logarithm of the *n*-octanol/water partition coefficient, and the molecular refractivity. In comparison with the results discussed in Ref. [9], the model gave more 2D-descriptors and thus may potentially be used in experimental studies. Five 3D-physicochemical properties including steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor factors play crucial roles in the binding affinities of inhibitors toward factor Xa [9]. Here, the SlogP descriptor ($P=C_{n\text{-octanol}}/C_{water}$; where $C_{n\text{-octanol}}$ and C_{water} are the concentrations of a solute in the lipid phase (*n*-octanol) and in the aqueous phase (water), respectively) relating to the absorption, transport, and excretion of drugs, i.e., the relative affinity for an aqueous (hydrophilic) or lipid (hydrophobic) medium, is present and contributes to pK_i. This could mean that the descriptor reflecting the hydrophobicity of the inhibitors is indispensable to the binding affinities toward factor Xa in 2D and 3D-QSAR studies. From the present results, the 2D-descriptors of b_ar, PEOE_VSA_POS, and SMR_VSA were found to contribute to pK_i. This result is helpful for further studies where these 2D-descriptors are not readily applicable. SMR_VSA5 and SlogP_VSA6 are descriptors based on the approximate accessible van der

Waals surface area (VSA), which is the surface area of a biomolecule that is accessible to a solvent, in unit of Å^2. Each atom has an accessible van der Waals surface area, ν_i, along with an atomic property, L_i. This property is in a specified range (a, b) and contributes to the descriptor. Thus, the SlogP_VSA6 is the sum of the ν_i from all atoms such that the L_i value of each atom, *i*, is in the range of (0.20, 0.25] [11]. The L_i contributes to the descriptor logP. The SMR_VSA5 refers to the sum of ν_i of all atoms such that the L_i value of each atom, *i*, is in the range of (0.44, 0.485) [12]. This L_i contributes to the descriptor the molecular refractivity (MR). The PEOE_VSA_POS denotes the sum of the van der Waals surface area of atom *i*, v_i, such that the partial charge of atom *i*, q_i, is non-negative. The atomic partial charges were calculated by partial equalization of orbital electronegativities (PEOE), in which charge is transferred between bonded atoms until equilibrium [13]. Descriptors using PEOE charges are prefixed with PEOE_. The positive coefficient signs of the descriptors represent a linear relationship between pK_i and the descriptors, i.e., the increase of these descriptors induces an increase in pK_i values (i.e., binding affinity decreases) while the negative coefficients imply an increase in binding affinity when the value of that descriptor increases.

The reliability of the developed model was evaluated via internal (cross), external, and total validations. The model gave statistically significant parameters for the external (12 inhibitors) and total (62 inhibitors) validations. The cross-validated squared correlation coefficient was Q^2_{cv}=0.789 and R^2=0.814, both of which are greater than 0.5. The RMSE values were lower than 0.5. These values confirmed the goodness of fit of the QSAR model. The model was also employed to predict the pK_i values of 62 inhibitors (Table 1).

By plotting the experimental pK_i vs. predicted pK_i, a linear correlation between experimental pK_i and predicted values was found. This linear relationship indicated that the model has a good predictive ability. Note that, the regression line was described as y=ax+b instead of y=x to simply illustrate a linear trend between the experimental pK_i vs. predicted pK_i, i.e., the experimentally large/small pK_i value (low/high activity) of one compound, the predicted value should be also large/small.

Conclusions

A 2D-QSAR model was established from 50 (*R*)-3-amidinophenylalanine inhibitors. The number of aromatic bonds, the positive van der Waals surface area, the logarithm of the *n*-octanol/water partition coefficients, and the molecular refractivity are crucial descriptors that contribute to pK_i values. The results demonstrated that the hydrophobicity, which has been reported in 3D-QSAR studies, plays a pivotal role on the affinities of inhibitors toward factor Xa. The model in this work gave significantly statistical parameters. The pK_i values of all inhibitors were predicted by employing the established 2D-QSAR model and there was a linear relationship between the experimental and predicted pK_i values. These results indicate the reliability of the model and could be helpful to develop drug candidates based on (*R*)-3-amidinophenylalanine derivatives.

ACKNOWLEDGEMENTS

The authors are thankful to Ho Chi Minh city University of Technology and Education for supporting websites to download the scientific articles.

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] E.W. Davie, K. Fujikawa, W. Kisiels (1991), "The coagulation cascade: initiation, maintenance, and regulation", *Biochemistry*, **30(43)**, pp.10364-10370.

[2] R.G. MacFarlane (1964), "An enzyme cascade in the blood clotting mechanism, and its function as a biochemical amplifier", *Nature*, **202**, pp.498-499.

[3] E.W. Davie, O.D. Ratnoff (1964), "Waterfall sequence for intrinsic blood clotting", *Science*, **145(3638)**, pp.1310-1312.

[4] B.Y. Zhu, R.M. Scarborough (2000), "Factor Xa inhibitors: recent advances in anticoagulant agents", *Annual Reports in Medicinal Chemistry*, **35**, pp.83-102.

[5] M.M. Mueller, S. Sperl, J. Stürzebecher, W. Bode, L. Moroder (2002), "(*R*)-3-amidinophenylalanine-derived inhibitors of factor Xa with a novel active-site binding mode", *Biological Chemistry*, **383(7-8)**, pp.1185-1191.

[6] S. Sperl, A. Bergner, J. Stürzebecher, V. Magdolen, W. Bode, L. Moroder (2000), "Urethanyl-3-amidinophenylalanine derivatives as inhibitors of factor Xa. X-ray crystal structure of a trypsin/inhibitor complex and modeling studies", *Biological Chemistry*, **381(4)**, pp.321-329.

[7] M. Dixon (1953), "The determination of enzyme inhibitor constants", *Biochemical Journal*, **55(1)**, pp.170-171.

[8] K. Roy, S. Kar, R.N. Das (2015), *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*, Elsevier.

[9] M. Böhm, J. Stürzebecher, G. Klebe (1999), "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", *Journal of Medicinal Chemistry*, **42(3)**, pp.458-477.

[10] R. Veerasamy1, H. Rajak, A. Jain, S. Sivadasan, C.P. Varghese, R.K. Agrawal (2011), "Validation of QSAR models - strategies and importance", *International Journal of Drug Design and Discovery*, **2(3)**, pp.511-519.

[11] C. Hansch, A. Leo, D. Hoekman (1995), *Exploring QSAR Vol. 2: Hydrophobic, Electronic and Steric Constants*, Washington.

[12] S.A. Wildman, G.M. Crippen (1999), "Prediction of physiochemical parameters by atomic contributions", *Journal of Chemical Information and Computer Sciences*, **39(5)**, pp.868-873.

[13] J. Gasteiger, M. Marsili (1980), "Iterative partial equalization of orbital electronegativity - a rapid access to atomic charges", *Tetrahedron*, **36(22)**, pp.3219-3228.