Hydrophobic Property of (*R*)-3 Amidinophenylalanine Inhibitors Contributes to their Inhibition Constants with Thrombin Enzyme

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

[1]Department of Chemistry, Faculty of
Sciences, Nong Lam University, Vietnam

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

Correspondence

Hoang Minh Hao, Department of
Chemical Technology, Faculty of
Chemical and Food Technology, Ho Chi
Minh City University of Technology and
Education, Vietnam

Email: haohm@hcmute.edu.vn

Check for updates

VNU-HCM Press

ABSTRACT

Introduction: Thrombin is the key enzyme of fibrin formation in the blood coagulation cascade. Thrombin is released by the hydrolysis of prothrombinase which is generated from factor Xa and factor Va in the presence of calcium ion and phospholipid. The inhibition of thrombin is of therapeutic interest in blood clot treatment. Currently, potent thrombin inhibitors of (*R*)-3-amidinophenylalanine, derived from benzamidine-containing amino acid, have been developed so far. In order to quantitatively express a relationship between chemical structures and inhibition constants (K_i with thrombin enzyme in a data set of (*R*)-3-amidinophenylalanine inhibitors), we developed a quantitative structure-activity relationship (QSAR) modeling from a group of 60 (*R*)-3-amidinophenylalanine inhibitors. **Methods**: A database containing chemical structures of 60 inhibitors and their K_i values was put into molecular operating environment (MOE) 2008.10 software, and the two-dimensional (2D) physicochemical descriptors were numerically calculated. After removing the irrelevant descriptors, a QSAR modeling was developed from the 2D-descriptors and K_i values by using the partial least squares (PLS) regression method. **Results**: The results showed that the hydrophobic property, reflected through *n*-octanol/water partition coefficient (P) of a drug molecule, contributes mainly to K_i values with thrombin. The statistic parameters that give the information about the goodness of fit of a 2D-QSAR model (such as squared correlation coefficient of $R^2 = 0.791$, root mean square error (RMSE) = 0.443, cross-validated $Q^2_{cv} = 0.762$, and cross-validated $RMSE_{cv} = 0.473$) were statistically obtained for a training set (60 inhibitors). The R^2 and RMSE values were obtained by using a developed model for the testing set (9 inhibitors) ; the total set has statistically significant parameters. Furthermore, the 2D-QSAR modeling was also applied to predict the K_i values of the 69 inhibitors. A linear relationship was found between the experimental and predicted pK_i values of the inhibitors. **Conclusion**: The results support the promising application of established 2D-QSAR modeling in the prediction and design of new (*R*)-3-amidinophenylalanine candidates in the pharmaceutical industry.

Key words: (R)-3-Amidinophenylalanine inhibitors, blood clot, thrombin, 2D-QSAR

INTRODUCTION

Fibrin clot formation is an important process that heals a wound and stops any unwanted bleeding. However, an abnormal clot in the bloodstream leads to pain and swelling because the blood gathers behind the clot. As a result, a heart attack can occur. There are pathways (mechanisms) which lead to fibrin formation. The intrinsic pathway was proposed in which fibrin formation resulted from a series of stepwise reactions involving only proteins circulating in blood as precursors or inactive forms[1–3]. Proteins were activated by proteolytic reactions and converted to thrombin. The intrinsic mechanism can be triggered when thrombin is generated, leading to the activation of factor XI[2]. The extrinsic pathway requires tissue factor VII in blood[2–5]. Initially, a complex including factor VII was formed via calcium ion dependent reaction and then converted factor VII to factor

VIIa (a: activated). The activation of many factors, including factor V, VIII, IX and X, in sequence results in the generation and release of thrombin. When thrombin is formed, it converts fibrinogen to fibrin by proteolysis. Finally, the cross-linking reactions were catalyzed by an activated factor XIIIa to form a very strong fibrin clot[2].

As discussed above, thrombin is a key enzyme in fibrin formation. Therefore, inhibitors selective toward thrombin have been developed; these include peptide aldehydes[6] and boronic acid derivatives[7]. The anticoagulants derived from 3-amidinophenylalanine that are associated with their inhibition constants (K_i values) toward thrombin enzyme have been reported[8,9]. The inhibition constant is an equilibrium constant of the reversible combination of the enzyme with a competitive inhibitor, I + E <->IE (K_i = [IE]/[I][E] ([I], [E] and [IE] are the equilibrium con-

centrations of inhibitor (I), enzyme (E), and enzyme-inhibitor complex (IE))[10]. The K_i value reflects the binding affinity of drug to target. The greater the binding affinity, the larger the K_i value is, *i.e.*, the less amount of medication needed to inhibit the enzyme. The design and synthesis of thrombin inhibitors could be improved in several ways. The two dimensional-quantitative structure-activity relationship (2D-QSAR) is one of the *in silico* drug discovery approaches due to its reliability and interpretability. In principle, the 2D-QSAR can be used to extract physicochemical properties (descriptors) which mainly contribute to the bioactivity of drug candidates[11]. In the present work, in order to express the 2D-descriptors playing a crucial role on K_i of a series of (*R*)-3-amidinophenylalanine inhibitors, we applied 2D-QSAR method to develop a mathematical QSAR equation from 60 inhibitors as a training set. The modeling was then used to predict K_i values of 69 inhibitors toward thrombin enzyme.

METHODS

A data set of 69 inhibitors derived from (*R*)- 3-amidinophenylalanine and their logarithm of inhibition constants, pK_i = - $logK_i$, toward thrombin enzyme was selected for the 2D-QSAR study[8] (**Figure 1**). Chemical structures of inhibitors were drawn in molecular operating environment (MOE) 2008.10 software and then optimized energetically prior to doing calculations. In order to develop a mathematical 2D-QSAR model, a training set containing 60 inhibitors was randomly selected in MOE. The selection of a training set was done when all parameters such as squared correlation coefficient (R^2), cross-validated correlation coefficient of Q^2_{cv}, and root-mean-square error (RMSE) of internal and external validations were statistically significant. In our study, this was repeated 8 times to obtain a satisfied training set. The remaining 9 inhibitors were used as a testing set to evaluate the reliability of the model. The input data were chemical structures and pK_i values of inhibitors. The 2D-molecular physicochemical properties (descriptors) are numerical values and calculated by using MOE. The inhibition constants, K_i, depended on 184 2D-molecular descriptors. However, the irrelevant descriptors which showed a zero value, a low correlation (< 0.07) with K_i, and high intercorrelation (> 0.7) between themselves were discarded. These descriptors were screened out using the Rapidminer 5 software. In addition, QuaSAR-Contigency and Principle Components in MOE 2008.10 were also used to screen the most relevant descriptors. The partial least squares (PLS) regression method was used

to develop a 2D-QSAR model. This model was used to predict the K_i values of 69 inhibitors and were predicted via the QuaSAR Fit validation panel in MOE.

RESULTS

2D-QSAR modeling

The first goal of this work is to develop a 2D-QSAR modeling which presents molecular descriptors of (*R*)-3-amidinophenylalanine inhibitors which predominantly contribute to the inhibition constant, K_i. The selected 2D-QSAR equation is given below:

$$pK_i = 5.774 - 2.458 \times SlogP\_VSA0 + 1.318 \times SlogP\_VSA1 + 1.559 \times SlogP\_VSA3 \quad (1)$$

Here, SlogP_VSA0, SlogP_VSA1, SlogP_VSA3 are molecular descriptors associated with coefficients. The training set was randomly selected, we have analyzed to develop significant models by using different training set with additional descriptors. The goal was to explain and search for other descriptors that relate to the inhibition constant. Unfortunately, other developed models possessed poor R^2, Q^2_{cv} and RMSE parameters. Therefore, those models could not be used for further analysis and discussion.

Statistical parameters

The statistical parameters (such as R^2, Q^2_{cv} and RMSE) give information about the goodness of fit of a model. The best model is selected when it possesses highest R^2 values, Q^2_{cv} (> 0.5) values, and lowest RMSE (< 0.5)[11]. **Table 1** shows the significantly statistical parameters of the internal, external (testing set), and total validations.

Predicted *p*K_i values using a developed 2D-QSAR model

Lastly, the pK_i values of 69 inhibitors were predicted using the established 2D-QSAR modeling. The pK_i values of all molecules are listed in **Figure 1.** A plot of experimental *vs*. predicted pK_i is shown in **Figure 2**.

Table 1: Statistically significant parameters of the established 2D-QSAR model

	Training set	Cross-validation	Testing set	
N^o	60	60	9	69
R^2	0.791		0.962	0.771
Q^2_{cv}		0.762		
RMS	0.443	0.473	0.161	0.460

Figure 1: Chemical structures, experimental (Exp) pK$_i$ [8,9] and predicted (Pred) pK$_i$ values toward thrombin of (*R*)-3-amidinophenylalanine inhibitors. R^1 and R^2 are the substituted groups in (*R*)-3-amidinophenylalanine skeleton
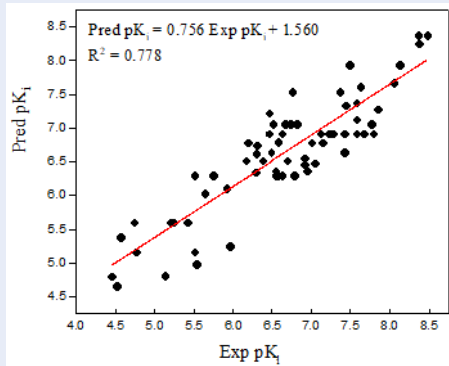
Figure 2: **The plot of correlations representing the experimental *vs.* predicted pK$_i$ values for 69 (*R*)-3-amidinophenylalanine inhibitors.**

DISCUSSION

Molecular descriptors

By using the partial least squares regression method, a 2D-QSAR modeling was established from a data composing of numerically relevant descriptors and pK$_i$ values of 60 inhibitors with thrombin. The developed modeling expressed the dependence of pK$_i$ on the hydrophobic descriptor. The logP refers to logarithm of the *n*-octanol/water partition coefficient (P). This property is an atomic contribution model that calculates logP from the given structure[12]. This descriptor was used as a measure of cell permeability of the drug molecule. The partition coefficient is a ratio between the concentrations of a solute in lipid phase (*n*-octanol) and in aqueous phase (P = C$_{n-octanol}$/C$_{aqueous}$). Compounds possessing P > 1 are lipophilic or hydrophobic while compounds for which P < 1 are hydrophilic. LogP of a molecule was calculated from fragmental or atomic contributions (surface area, molecular properties, and solvatochromic parameters) and various correction factors (electronic, steric, or hydrogen-bonding effects)[11,13]. Each atom has an accessible van der Waals surface area (VSA), a$_i$, along with an atomic property, p$_i$. This property is in a specified range (a, b) and contributes to the descriptor. Slog P_VSA is the sum of a$_i$ of all atoms, such that p$_i$ value of each atom i is in a range of (a, b) (**Table 2**) ; *p$_i$* contributes to descriptor logP[13]. The sign and magnitude of the descriptors coefficients re present the contribution of each descriptor to pK$_i$. Positive coefficients imply that pK$_i$ values of molecules increase with increasing SlogP_VSA values, while negative values demonstrate an increase in

pK$_i$ (*i.e.*, K$_i$ -binding affinity decreases) with decreasing values of the descriptors. The higher the absolute coefficient value is, the more crucial the contribution of the descriptor on the binding affinity. The modeling indicates that inhibitors possessing higher SlogP_VSA1 and SlogP_VSA3 properties will result in a decrease in K$_i$ values, *i.e.*, binding affinities decrease while an increase in SlogP_VSA0 property would induce a better binding affinity.

Table 2: **Molecular descriptors in 2D-QSAR modeling**

Descriptor Code	Description
SlogP_VSA0	Sum of a_i such that p_i <= -0.4
SlogP_VSA1	Sum of a_i such that p_i is in (-0.4, -0.2]
SlogP_VSA3	Sum of a_i such that p_i is in (0.0, 0.1]

2D-QSAR modeling and its validation

The selected 2D-QSAR modeling is a model possessing statistically significant parameters of internal and external validations. The developed model (**Equation (1)**) from the training set has showed R^2 value of 0.791 and RMSE value of 0.443. These values confirmed the reliability of the model. As mentioned, the reliability and statistical relevance of the 2D-QSAR modeling was examined by internal and external validation procedures. Internal validation was applied by Leave One Out (LOO) cross-validation (CV)[11,14]. The values of Q$^2_{cv}$ > 0.5 and RMSE< 0.5 (**Table 1**) further supported the reliability and interpretability of the modeling. The pK$_i$ values of inhibitors were predicted by applying an established 2D-QSAR modeling on a total set. By plotting the predicted pK$_i$ values vs. the experimental ones (**Figure 2**), there is a linear relationship between the predicted and experimental pK$_i$ values of inhibitors, *i.e.*, both pK$_i$ values are high (a low inhibitory activity) or low (a good inhibitory activity). These results show that the modeling is reliable to predict the pK$_i$ values of the inhibitors.

CONCLUSIONS

The 2D-QSAR modeling has been successfully developed from 2D-descriptors of 60 (*R*)-3-amidinophenylalanine inhibitors associated with their inhibition constants, K$_i$. The established QSAR modeling was internally, externally, and totally validated, demonstrating satisfactory statistical parameters. Hydrophobicity is an important descriptor

in the modelling of binding affinity. The 2D-QSAR equation was applied to predict K_i values of all inhibitors. The results revealed a good predictability of the modeling. Based on the developed 2D-QSAR modeling, the design of the new inhibitors derived from (*R*)-3-amidinophenylalanine should focus on the hydrophobicity of derivatives by theoretical calculations to obtain the numerical values of hydrophobic descriptors. The chemical structures of inhibitors possessing lower values of SlogP_VSA1, SlogP_VSA3 descriptors and higher SlogP_VSA0 descriptor should be further studied in synthetic experiments.

LIST OF ABBREVIATIONS

2D-QSAR: two dimensional-quantitative structure-activity relationship
CV: cross-validation
LOO: leave one out
MOE: molecular operating environment
RMSE: root-mean-square error

AUTHOR CONTRIBUTIONS

The contributions of all authors are equal in selecting a data, calculating descriptors, analyzing results and writing a manuscript.

COMPETING INTERESTS

The authors declare that they have no competing interests.

ACKNOWLEDGMENT

REFERENCES

1. Davie EW, Ratnoff OD. Waterfall sequence for intrinsic blood clotting. Science [Internet]. 1964 Sep 18;145(3638):1310–1312. [cited 2019 May 22]. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.145.3638.1310.
2. Davie EW, Fujikawa K, Kisiel W. The coagulation cascade: initiation, maintenance, and regulation. Biochemistry (Mosc) [Internet]. 1991 Oct;30(43):10363–70. [cited 2019 May 22]. Available from: http://pubs.acs.org/doi/abs/10.1021/bi00107a001.
3. Maynard JR, Heckman CA, Pitlick FA, Nemerson Y. Association of tissue factor activity with the surface of cultured cells. J Clin Invest [Internet]. 1975 Apr 1;55(4):814–838. [cited 2019 May 22]. Available from: http://www.jci.org/articles/view/107992.
4. Bach R, Nemersonl Y, Konigsber W. Purification and Characterization of Bovine Tissue Factor;256(16):8324–31.
5. Broze GJ. Binding of human factor VII and VIIa to monocytes. J Clin Invest [Internet]. 1982 Sep 1;70(3):526–35. [cited 2019 May 24]. Available from: http://www.jci.org/articles/view/110644.
6. Bagdy D, Barabs E, Szab G, Bajusz S, Szll E. In vivo anticoagulant and antiplatelet effect of D-Phe-Pro-Arg-H and D-MePhe-Pro-Arg-H. Thromb Haemost. 1992 Mar 2;67(3):357–65.
7. Hussain MA, Knabb R, Aungst BJ, Kettner C. Anticoagulant activity of a peptide boronic acid thrombin inhibitor by various routes of administration in rats. Peptides. 1991 Oct;12(5):1153–4.
8. Böhm M, Stürzebecher J, Klebe G. Three-Dimensional Quantitative StructureActivity 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 [Internet]. 1999 Feb;42(3):458–77. [cited 2019 May 27]. Available from: https://pubs.acs.org/doi/10.1021/jm981062r.
9. Stürzebecher J, Prasa D, Hauptmann J, Vieweg H, Wikström P. Synthesis and StructureActivity Relationships of Potent Thrombin Inhibitors: Piperazides of 3-Amidinophenylalanine. J Med Chem. 1997 Sep;40(19):3091–9.
10. Dixon M. The determination of enzyme inhibitor constants. Biochem J [Internet]. 1953 Aug;55(1):170–1. [cited 2019 May 24]. Available from: http://www.biochemj.org/cgi/doi/10.1042/bj0550170.
11. Roy K, Kar S, Das RN. Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment; 2015.
12. Wildman SA, Crippen GM. Prediction of Physicochemical Parameters by Atomic Contributions. J Chem Inf Comput Sci [Internet]. 1999 Sep 27;39(5):868–73. [cited 2019 May 27]. Available from: Availablefrom:https://pubs.acs.org/doi/10.1021/ci990307l.
13. Martin YC. Exploring QSAR: Hydrophobic, Electronic, and Steric Constants C. Hansch, A. Leo, and D. Hoekman. American Chemical Society, Washington, DC. 1995. Xix + 348 pp. 22 × 28.5 cm. Exploring QSAR: Fundamentals and Applications in Chemistry and Biology. C. Hansch and A. Leo. American Chemical Society, Washington, DC. 1995. Xvii + 557 pp. 18.5 × 26 cm. ISBN 0-8412-2993-7 (set). $99.95 (set). J Med Chem [Internet]. 1996 Jan [cited 2019 May 27];39(5):1189–90; 1995. Available from: https://pubs.acs.org/doi/10.1021/jm950902o.
14. Ghose AK, Viswanadhan VN, editors. Combinatorial library design and evaluation: principles, software tools, and applications in drug discovery. New York: M. Dekker. New York: M. Dekker; 2001.