Chronic Diseases Prevention Review 2 (2017) 32-36

Available at http:// www.cancercellresearch.org ISSN 2158-0820

# QSAR Study of Mer Specific tyrosine kinase inhibitors for the treatment and prevention of thrombosis

Meng Li, Wensi Wu, Xiaotong Liu, Linhao Xu*

Department of Thoracic Surgery, the Affiliated Hospital of Qingdao University, Qingdao, Shandong 266003, China

***Abstract:*** **In this work, a quantitative structure–activity relationship (QSAR) model was used to predict the Mer Specific tyrosine kinase inhibitors by the gene expression programming (GEP). A data set of 43 compounds was used. First, we used the HM to select the descriptors and set up linear models. The result shows that $R^2$=0.7140, F=14.98 $S^2$=2.1697 .Then the data set was divided into two subsets of training and testing. Gene Expression Programming (GEP) was applied to select the respective variables to build nonlinear QSAR models. The GEP method produced a nonlinear and five descriptor quantitative models with a mean error and a correlation coefficient of 0.173 and 0.9549 for the training set, 6.499 and 0.774 for the test set, respectively. Thus the QSAR model will have a better evaluation in providing guidance for designing and synthesizing of anti-thrombosis drugs in the future.**

***Keywords:*** QSAR; Gene expression program; Mer Specific tyrosine kinase inhibitors; Heuristic method

Received 19 February 2017, Revised 23 April 2017, Accepted 25 April 2017

*Corresponding Author: Linhao Xu, qdqyfyxulinhao@126.com

## 1. Introduction

Pathologic thrombus formation may result in stroke or heart attack. Therefore, Anti-platelet compounds are an important family of drugs for cardiovascular diseases and for certain surgical procedures where a risk of stroke or thrombosis. Mer is a member of the TAM (Tyro3, Axl and Mer) receptor tyrosine kinase (RTK) subfamily with growth-arrest-specific-6Gas6) as one of the endogenous ligands [1,2]. Recently, Mer has been shown to play important roles in regulating macrophage activity and platelet aggregation.

The role of Mer kinase in regulating the second phase of platelet activation generates an opportunity to use Mer inhibitors for preventing thrombosis with diminished likelihood for bleeding as compared to current therapies. QSAR studies identify compound 23 (UNC2881) as a lead compound for in vivo evaluation. When applied to live cells, 23 inhibitors steady-state Mer kinase phosphorylation with an $IC_{50}$ value of 22nM. Treatment with 23 is also sufficient to block EGF-mediated stimulation of a chimeric receptor containing the intracellular domain of Mer fused to the extracellular domain of EGFR. In addition, 23 potently inhibitors collagen-induced platelet aggregation suggest that this class of inhibitors may have utility for prevention and/or treatment thrombosis.

Designing compounds possessing Mer Specific tyrosine kinase inhibitory activities can be helpful to teat such cases; however we are not always sure about the biological activities of designed novel compounds. Several experiments should be run to synthesize, and then tested to illustrate the biological impacts of the newly designed molecules, however we are faced with time limitation and high cost of experimental runs. Consequently, it is of interest to develop a model to predict the biological activities before any experimental works for synthesis [3].

There has been growing interest over computational methods to predict the biological activity by chemical structure, so as to decide whether it has objective qualities or not. In this contribution, the well-known method, which is called quantitative structure activity relationship has been developed. The QSAR (Quantitative structure activity relationship) tries to explain the observed variance in the biological effect of certain classes of compounds as a function of molecular changes caused by the substituents. These physiochemical descriptors which included parameters to account for hydrophobicity, electronic properties and steric effect are determined empirically or, more recently, by computational methods [4]. QSAR methodologies save resources and expedite the process of the development of new molecules and drugs [5-7].

The GEP (Gene Expression Programming) is a genotype/phenotype system that involves computational programs with different sizes and shapes of expression trees. They encoded the chromosomes of the fixed length. It has been successfully used to build QSAR model [8,9] to predict evaporation estimation [10], cement strength [11] and milling surface roughness [12]. In our works, we calculated many descriptors from the 43 molecular structures in literature1 by the software CODESSA. Heuristic method (HM) was used to select descriptor and set up linear model. Then we developed a new QSAR model to explore the $IC_{50}$ of anti-thrombosis the compounds with diverse structures. It will show that the GEP prediction have a good result.

## 2. Material and methods

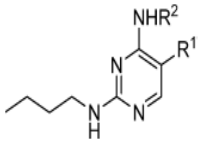
### *2.1. Data set*

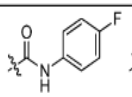
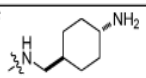
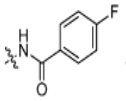
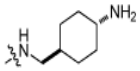
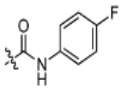
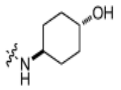
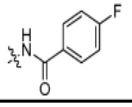
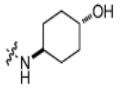
In this study, a dataset comprising 43 compounds of Mer Specific tyrosine kinase inhibitors were taken

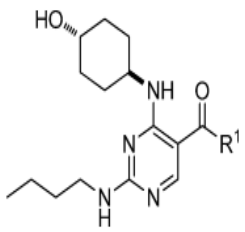
 

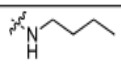
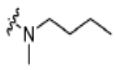
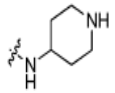
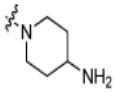
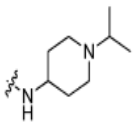
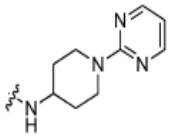
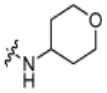
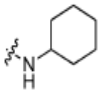
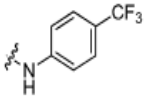
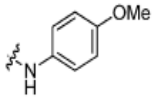
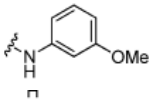
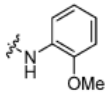
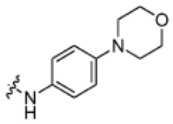
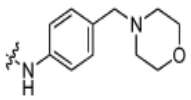
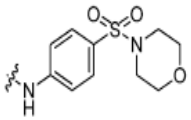
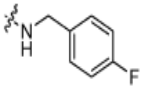
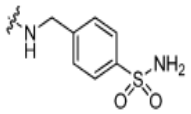
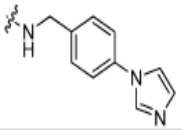
from the literature [1]. The chemical structure and activity data of the molecules are listed in Table 1. The data set was divided into two training and tests sets in which 30 and 13 molecules were obtained for each subset respectively. The test sets are used to assess the predictive ability of model which was obtained based on training set.

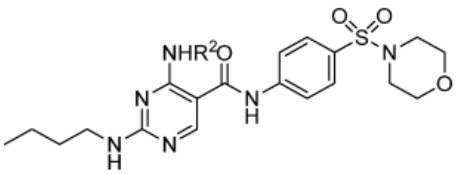
**Table 1. Key SAR Observations**



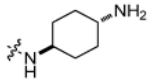
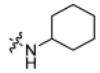
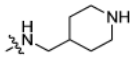
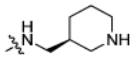
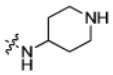
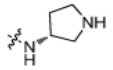
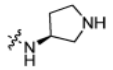
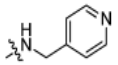

| Compound | $R^1$ | $NHR^2$ | $IC_{50}$ (μM)[a] | | |
|---|---|---|---|---|---|
| | | | Mer | Axl | Tyro3 |
| 2 |  |  | 0.040 | 2.0 | 0.90 |
| 3 |  |  | 12.6 | >30 | >30 |
| 4 |  |  | 0.023 | 1.7 | 2.2 |
| 5 |  |  | >30 | >30 | >30 |

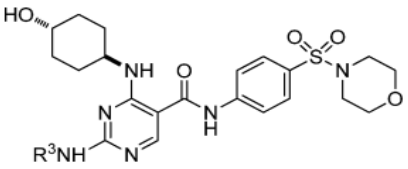


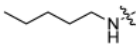
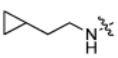
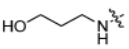
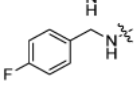

| Compound | $R^1$ | $IC_{50}$ (μM)[a] | | |
|---|---|---|---|---|
| | | Mer | Axl | Tyro3 |
| 6 |  | 0.067 | 3.0 | 6.7 |
| 7 |  | 5.9 | >30 | >30 |
| 8 |  | 0.0090 | 0.48 | 0.53 |
| 9 |  | 2.5 | >30 | >30 |
| 10 |  | 0.0084 | 0.43 | 0.19 |
| 11 |  | 0.0041 | 0.56 | 0.48 |
| 12 |  | 0.013 | 0.58 | 0.54 |
| 13 |  | 0.0080 | 0.37 | 0.37 |
| 14 |  | 0.23 | 16.6 | 17.9 |
| 15 |  | 0.015 | 0.88 | 0.66 |
| 16 |  | 0.029 | 1.7 | 0.83 |
| 17 |  | 0.039 | 3.8 | 1.5 |
| 18 |  | 0.014 | 0.57 | 0.45 |
| 19 |  | 0.0067 | 0.37 | 0.24 |
| 20 |  | 0.0063 | 0.53 | 0.70 |
| 21 |  | 0.014 | 0.79 | 0.57 |
| 22 |  | 0.0052 | 0.26 | 0.28 |
| 23 |  | 0.0043 | 0.36 | 0.25 |



| Compound | $NHR^2$ | $IC_{50}$ (μM)[a] | |
|---|---|---|---|
| | | Mer | Axl |
| **24** |  | 0.0021 | 0.40 |
| **25** |  | 0.59 | >30 |
| **26** |  | 0.037 | 4.7 |
| **27** |  | 0.14 | 14.0 |
| **28** |  | 0.062 | 10.7 |
| **29** |  | 1.0 | >30 |
| **30** |  | 0.11 | 12.6 |
| **31** |  | 0.62 | 21 |
| **32** |  | 0.17 | >30 |
| **33** |  | 9.9 | >30 |



| Compound | $NHR^3$ | $IC_{50}$ (μM)[a] | |
|---|---|---|---|
| | | Mer | Axl |
| **34** |  | 2.4 | 16.6 |
| **35** |  | 0.15 | 1.2 |
| **36** |  | 0.037 | 0.78 |
| **37** |  | 0.011 | 0.61 |
| **38** |  | 0.027 | 1.1 |
| **39** |  | 0.0073 | 0.30 |
| **41** |  | 0.059 | 4.3 |
| **42** |  | 0.024 | 1.9 |
| **43** |  | 0.047 | 9.0 |
| **44** |  | 0.18 | 7.4 |
| **45** |  | 0.12 | >30 |

### *2.2. Calculation of the descriptors*

Firstly, we should draw the molecular structural formula by Chem Draw. Then, the structures of compounds were first pre-optimized with the Molecular Mechanics Force Field (MM+) procedure included in Hyper Chem. It can get a more precise optimization by the semi-empirical AM1 method in MOPAC [13]. The calculated descriptors were analyzed to remove the constant and near constant variables in order to reduce the redundancy existing in descriptor data matrix. Correlation of remained descriptors and inhibition activity was examined to omit the collinear descriptors (i.e. r> 0.90). Finally, there are some descriptors remained. Then Split the data into training and test set. The molecular structures were optimized using the Polak–Ribiere algorithm till the root mean square gradient was 0.01. The MOPAC output files were used in CODESSA program to calculate classes of descriptors. The molecular descriptors are divided into six groups: constitutional, topological, geometrical, electrostatic, quantum-chemical and thermodynamic [14].

### *2.3. GEP theory*

Gene Expression Programming (GEP) [15] is a learning algorithm and it was invented by ferreira in 1999. It learns specifically about relationships between variables in sets of data and then builds models to explain these relationships. The GEP system is a full-fledged genotype/phenotype system with expression trees of different sizes and shapes encoded in linear chromosomes of fixed length. Also the most important things are that GEP chromosomes are multi-genetic, encoding multiple expression trees or sub-programs that can be organized into a much more complex program. So, like the DNA/protein system of life on Earth, the gene/tree system of GEP can not only explore all the crannies and paths of the solution space but it's also free to explore higher levels of organization.

## 3. Results and discussion

### *3.1. Result of HM*

Molecular descriptors were produced. Then Heuristic method (HM) was used to select descriptor and set up linear model. (1) There is no limit. (2) Select several relevant descriptors to statistical data. The quality of the correlation was tested by the $R^2$, $R_{CV}^2$, F, and the $S^2$. With the growing of number of descriptors $R^2$, $R_{CV}^2$ and$S^2$ produced corresponding changes. While selecting 6 descriptors, the values of $R^2$, $R^2$cv and $S^2$ tend to be stable. The corresponding physicochemical meaning and statistical parameters are in the Table 2.

IC50= -8.44 MAOEP+23.29 IOTOKSE+ 2.17 MB00AOA-0.47 TPCCOMD+24.59 ABOOACA+28.55 MNRIFAOA (1)

$R^2$=0.714 $R_{CV}^2$=0.77 $S^2$=2.1697 F=14.98

**Table 2 Corresponding physicochemical meaning and statistical parameters**

| | Physical-chemical meaning | Coefficient | Standard error | T-test |
|---|---|---|---|---|
| MAOEP | Min atomic orbital electronic population | -8.44 | 2.62 | -3.23 |
| IOTOKSE | Image of the Onsager-Kirkwood solvation energy | 23.29 | 4.77 | 4.89 |
| MBOOAOA | Max bond order of a O atom | 2.17 | 0.68 | 3.19 |
| TPCCOMD | Tot point-charge comp. of the molecular dipole | -0.47 | 0.16 | -2.85 |
| ABOOACA | Avg bond order of a C atom | 24.59 | 8.63 | 2.85 |
| MNRIO | Min nucleoph react index for a O atom | 2855 | 1221 | 2.34 |

### *3.2. Results of GEP*

We used the software automatic problem solver (APS) to predict $IC_{50}$ of Mer Specific tyrosine kinase inhibitors because it allows an easy optimization of intermediate solutions and easy test of the evolved models against a test set. The function set is expressed as F= {'+', '-', '*', '/', 'sin', 'cos', 'ln',}. A small population of 43 individuals is calculated by [19]

$$E_i = \frac{1}{n}\sum (P_{(ij)} - T_j)^2 \quad (2)$$

In the run, the 299 generation is a very good solution, with a $R^2$ of 0.95 for the training set and a $R^2$ of 0.77 for the testing set. The expression tree can be converted into Eq. (3).

$$Y = (x2 \times x3 / x5) \div ((x1 - x3) + \log x4) + (\sin(\cos(\log x2)) \div (x4 \times x5) + ((x1 \div (x5 \times \log x2) + x5) + (x1 / x3) \quad (3)$$

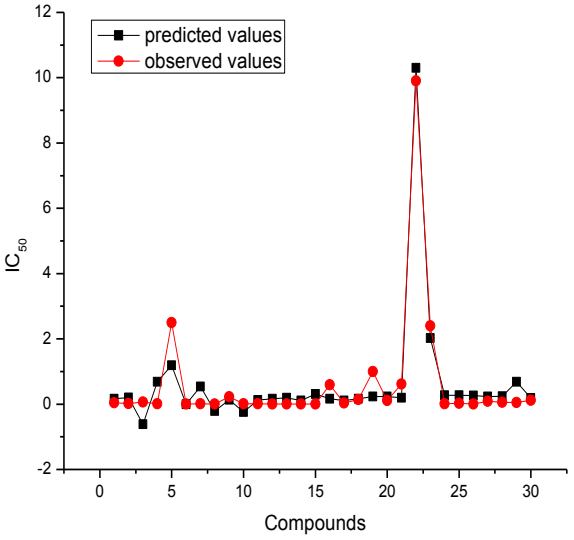



**Figure 1. Fitting curve of experimental and calculated IC50 in training set by GEP.**

In the Eq. (6) X1, X2, X3, X4, X5, X6 represent variables of MAOEP, IOTOKSE, MBOOAOA, TPCCOMD, ABOOACA, MNRIFAOA. GEP model can accurately predict the activity of Mer Specific Tyrosine Kinase Inhibitors with the error less than 0.1. The model for most of the Mer Specific Tyrosine Kinase Inhibitors predicted values are lower than the experimental value. It may be related to MAOEP and TPCCOMD has excessive negative effect.

The fitting situation of the training and the test set for the GEP model are shown in Figure 1 and 2 respectively.

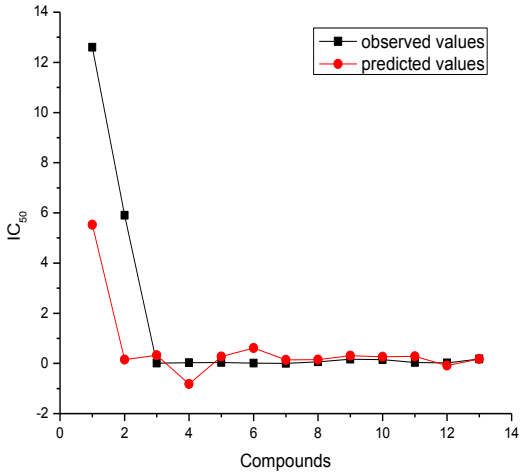



**Figure 2. Fitting curve of experimental and calculated IC50 in testing set by GEP.**

**Table 3 Parameters for the simple symbolic regression problem.**

| Parameter names | Values |
|---|---|
| Number of chromosomes | 100 |
| Function sets | "+""—""×" |
| Gene head size | "÷""cos""sin"" ln" |
| Number of genes | 8 |
| Linking function | 5 |
| Mutation rate | + |
| Inversion rate | 0.044 |
| one-point recombination rate | 0.1 |
| two- point recombination rate | 0.3 |
| Gene recombination | 0.3 |
| Gene transposition | 0.1 |
| IS transposition rate | 0.1 |
| RIS transposition rate | 0.1 |
| RIS elements' length | 0.1 |

| | |
|---|---|
| Gene transposition rate | 100 |
| Selection range | 0.01 |
| Precision | 100 |
| | 0.1 |

### *3.3. The GEP compares with HM*

In the GEP method, $R^2$ of the training set 0.949 and the test set 0.774 are all greater than $R^2$ of HM 0716. The results of GEP method are better than that of HM method. SO, the GEP model has a better prediction of the log ($IC_{50}$) than HM.

From the above discussion, it is obvious that the physical chemistry property of compounds influences significantly the value of $IC_{50}$. As can be seen from the above, the 6 molecular descriptors are the most important physical and chemical parameters to predict the values of $IC_{50}$. What is more, MNRIFAOA has the most effective on the activity of Mer Specific Tyrosine Kinase Inhibitors but THE MAOEP has the most negative effect.

## 4. Conclusion

In this work, the quantitative structure–activity relationship study of a series of Mer Specific tyrosine kinase inhibitors was carried out. Different kinds of descriptors were calculated to represent the biological activity through the molecular structure. To select the respective descriptors, Polak–Ribiere algorithm was used and provided the five main descriptors available for the presenting biological activities. For splitting the data into two subsets, training set and testing set. In this study, we can draw a conclusion that GEP has a good prediction of drug .By this means, we can not only save some money and times but also we can predict the expected activity of drugs and aiding in drug design.

## References

[1] dx.doi.org/10.1021/jm4013888 | J. Med. Chem. 2013, 56:9693-970

[2] Laurance S, Lemarie CA, Blostein MD. Growth Arrest-Specific Gene 6 (gas6) and Vascular Hemostasis. Adv Nutr, 2012, 3(2):196-203.

[3] Fujita T. Qsar and Drug Design: New Development and Applications, Elsevier, Amsterdam, Netherlands, 1995, 1:413-450.

[4] Sahu V, Sharma P, Kumar A. Synthesis and QSAR modeling of 1-[3-methyl-2-(aryldiazenyl)-2H-aziren-2-yl]et hanones as potential antibacterial agents. Med. Chem. Res., 2013, 22(5):2476-2485.

[5] Bhadoriya KS, Kumawat NK, Bhavthankar SV, et al. Exploring 2D and 3D QSARs of benzimidazole derivatives as transient receptor potential melastatin 8 (TRPM8) antagonists using MLR and kNN-MFA methodology, J. Saudi Chem. Soc., 2012, 22:13-23.

[6] Bhadoriya KS, Sharma MC, Jain SV. Pharmacophore modeling and atom-based 3D-QSAR studies on amino derivatives of indole as potent isoprenylcysteine carboxyl methyltransferase (Icmt) inhibitors, J. Mol. Struc., 2014, 11:222-228.

[7] Bhadoriya KS, Sharma MC, Jain SV, et al. QSAR Studies of Fused 5, 6-Bicyclic Heterocycles as γ-Secretase Modulators, J. Pharm. Res., 2012, 5(8): 4127–4132.]

[8] Ozlem T. Evaporation estimation using gene expression programming, Journal of Applied Sciences, 2005, 5:508-512.

[9] J.P.P. Stewart, MOPAC 6.0; Quantum Chemistry Program Exchange;QCPE, No. 455.

[10] Wright JS, Carpenter DJ, McKay DJ. Theoretical calculation of substituent effects on the O–H bond dissociation enthalpy, Journal of the American Chemical Society, 1985, 107:3902-3909.

[11] Ferreira C. Gene expression programming: a new adaptive algorithm for solving problems, Complex Systems, 2001, 13:87-129.

[12] Sannigrahi AB. Advances in Quantum Chemistry, 1992, 23:301-351.

[13] Clementi E. Computational Aspects of Large Chemical Systems, Springer Verlag, New York, 1980.

[14] http://www.gepsoft.com/gepsoft

[15] http://www.gene-expression-programming.com.