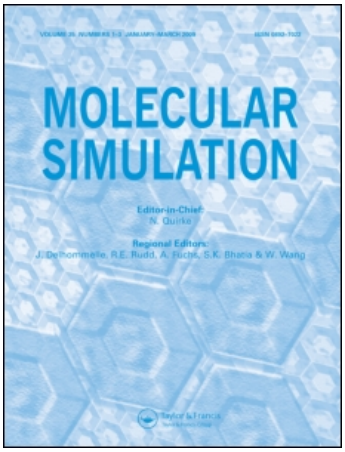


Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

QSAR study of substituted 1-(2-naphthyl)-1H-pyrazole-5-carboxylamide factor Xa (fXa) inhibitors

X. Ma[a]; L. Zhou[a]
[a] Department of Applied Chemistry, College of Chemical Engineering, Sichuan University, Sichuan, Chengdu, P. R. China

Online publication date: 27 July 2010

To cite this Article Ma, X. and Zhou, L.(2007) 'QSAR study of substituted 1-(2-naphthyl)-1H-pyrazole-5-carboxylamide factor Xa (fXa) inhibitors', Molecular Simulation, 33: 15, 1213 — 1222
To link to this Article: DOI: 10.1080/08927020701730351
URL: http://dx.doi.org/10.1080/08927020701730351

Molecular Simulation, Vol. 33, No. 15, December 2007, 1213–1222

Taylor & Francis
Taylor & Francis Group

QSAR study of substituted 1-(2-naphthyl)-1H-pyrazole-5-carboxylamide factor Xa (fXa) inhibitors

X. MA and L. ZHOU*

Department of Applied Chemistry, College of Chemical Engineering, Sichuan University, Sichuan, Chengdu 610065, P. R. China

(Received June 2007; in final form October 2007)

Blood coagulation factor Xa (fXa), a trypsin-like serine protease that plays a pivotal role in the blood coagulation cascade, has emerged as a very attractive target for the design of new therapeutic agents with potential for the treatment of arterial and venous thrombosis. Two-dimension quantitative structure–activity relationship studys such as partial least squares, artificial neural networks (ANN), genetic algorithm optimized ANN (GA-NN) have been carried out for some potent 1-(2-naphthyl)-1H-pyrazole-5-carboxylamides fXa inhibitors. The results showed that the nonlinear models have better predictive abilities than linear model. The GA-NN model showed the best performance.

Keywords: factor Xa inhibitors; Partial least squares (PLS); Artificial neural networks (ANN); Genetic algorithm optimized ANN (GA-NN); QSAR

1. Introduction

Thrombosis-related ischemic diseases are leading causes of death in the world. Unfortunately, these diseases are still treated by relatively antiquated drugs. The orally bioavailable anticoagulants on the market for thrombotic disorders, suffer from a number of shortcomings which limit its application. However, an exciting new wave of antithrombotic compounds has recently emerged in clinical trials. A particularly attractive new class of antithrombotic agents is the direct factor Xa (fXa) inhibitors, which appear to provide an enhanced risk-to-benefit margin compared to conventional anticoagulation therapies [1,2].

fXa, a trypsin-like serine protease that plays a pivotal role in the blood coagulation cascade. It situated at the convergence of the surface-activated intrinsic and factor-activated extrinsic coagulation pathways. The prothrombinase complex is formed by fXa on the phospholipid surface with factor Va and calcium; it catalyzes the proteolysis of prothrombin to thrombin (factor IIa) [3]. Thrombin is the main, final enzyme in the phospholipid coagulation system that leads to fibrin formation. It provides positive and negative feedback regulatory signal in the normal hemostasis, while in pathological conditions fXa provides catalytic activation of thrombin. Thus, the inhibition fXa affects the coagulation but not the platelet function. As suggested by preclinical studies, fXa inhibitors have less potential for increasing the risk of abnormal bleeding and have a wider therapeutic index. Representing an important advance in the management of both arterial and venous thrombosis, fXa inhibitors have been actively pursued as new orally active antithrombotic agents. Recently, the inhibition fXa has been intensely investigated in order to replace the existing therapies in the treatment or prevention of thromboembolic disorders [4–6].

Quantitative structure–activity relationship (QSAR) models are widely used in drug design and medicinal chemistry [7]. Each kind of model relates the chemical structures to their own biological activity through linear or nonlinear mathematical equation. Multiple linear regressions (MLRs) and partial least squares (PLS) are the typical linear models in QSAR researchs, and artificial neural networks (ANN) is a kind of nonlinear QSAR models. Different model has different prediction ability. In many areas, genetic algorithm (GA) is a general evolutionary algorithm that can be used for optimization [8]. GA can be used with neural networks in order to optimize the networks. In general, we can get a better model than solitary ANN.

In this paper, we used the GA to optimize the ANN. We refer this model as GA-NN. In order to study the relationship between the descriptors and the activities

*Corresponding author. Tel.: + 86-28-81938747. Email: zhouluscu@163.com

Molecular Simulation
ISSN 0892-7022 print/ISSN 1029-0435 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/08927020701730351

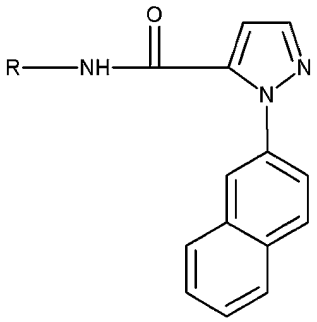


Figure 1. Structure of substituted1-(2-naphthyl)-1H-pyrazole-5-carboxylamide.

of the substituted1-(2-naphthyl)-1H-pyrazole-5-carboxylamide (figure 1) fXa inhibitors, PLS, ANN, GA-NN were used to build the QSAR models, and we compared the results of each model.

2. Material and methods

2.1 Experimental data

Sixty-seven substituted 1-(2-naphthyl)-1H-pyrazole-5-carboxylamide fXa inhibitors were taken from the studies reported by Zhaozhong *et al.* [4,9,10]. The basic structure of these compounds is shown in figure 2. The *in vitro* anticoagulant activities of these compounds were given in IC_{50} (nM) pattern. These values were converted to negative logarithm $\log(1/IC_{50})$ (briefly described as PIC50) as the dependent variable representing the biological activity of these compounds. The substituent patterns (SPs) of the compounds and the PIC_{50} values are given in table 1. The molecules were divided into three sets randomly: training set, validation set, and testing set. The training set (from number 1 to 55 in table 1) was used to build up the model and the testing set (from number 62 to 67 in table 1) was used to test the prediction ability of the model. The validation sets (from number 56 to 61 in table 1) was used during the development of ANN model and GA-NN model to prevent neural network overfitting.

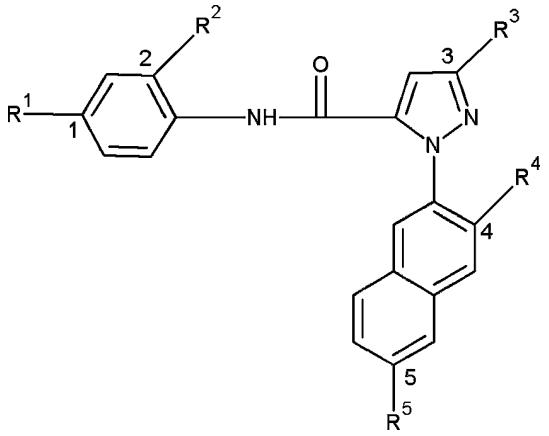


Figure 2. Basic structure of compounds.

2.2 Moleculars modeling and descriptors calculation

All of the molecular structures were drawn by HyperChem 7.0 for Windows (Hypercube, FL, USA). First, MM + molecular mechanics force field was selected to optimize geometrical conformations. After getting the optimized conformations, they were subjected to a refined geometry optimization using AM1 semiempirical molecular orbital theory. We selected the AM1 Hamiltonian as the optimization function because it gives good estimates of molecular energies and the computation time is much shorter than needed by *ab initio* methods. Based on the most optimized conformations, the physicochemical, geometrical and electronical descriptors of each molecule were calculated. All the 22 descriptors are summarized in table 2.

2.3 Partial least squares

Partial least square regression [11,12], a popular linear modeling technique, has been extensively used in QSAR studies, it is a method for relating dependent variables and independent variables by a linear multivariate model. In the typical setting, given a finite training set with n samples (x_i, y_i), PLS builds a linear relationship between x and y that is then used for prediction of y for new data x. The main assumption of PLS is that the data x, although possibly residing in a high-dimensional space, depend linearly on only a small number of latent variables. PLS estimates these latent variables as projections of the original input variables of x and uses them to construct the regression vector relating x to y.

In order to model the structure activity relationships of the compounds, PLS regression was employed in this study. The PLS regression was performed by Minitab (version 15, MINITAB, Inc.), A leave-one-out cross validation procedure was used to obtain the best PLS model. The model refinement procedure uses the predicted residual errors sum of squares (PRESS) of the cross validation to select the least or the most significant variables and also the optimum number of latent variables [13]. After calculation of PRESS in optimum number of factors, a new descriptor was entered and the PLS modeling was run again. If the PRESS at this step was lower than the former one, the descriptor was entered in the model; Otherwise, the descriptor was removed. The predicted values for the test sets can be compared to the experimental values by calculating root-mean-square error (RMSE). PRESS is defined as:

$$\mathrm{PRESS} = \sum_{i=1}^{n} (y_i - y)^2 \tag{1}$$

Where y_i is the calculated value of the ith object and y is the corresponding experimental value of this object. RMSE is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{\mathrm{PRESS}}{n}} \tag{2}$$

Where n is the number of compounds.

Table 1. Substituent patterns (SPs) and PIC_{50} for basic structure.

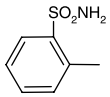
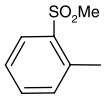
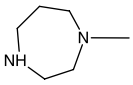
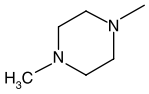
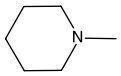
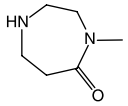
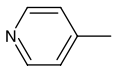
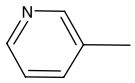
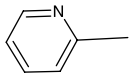
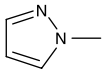
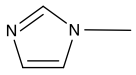
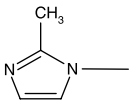
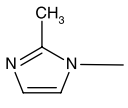
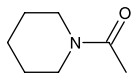
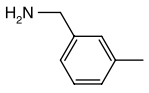
Number	R_1	R_2	R_3	R_4	R_5	PIC_{50}
1		—H	—CH_3	—H	—Cl	7.95861
2		—H	—CH_3	—H	—Cl	8.39794
3		—H	—CH_3	—H	—Cl	7.4437
4		—H	—CH_3	—H	—Cl	6.77989
5		—H	—CH_3	—H	—Cl	6.88273
6		—H	—CH_3	—H	—Cl	6.88606
7		—F	—CH_3	—H	—Cl	7.39794
8		—F	—CH_3	—H	—Cl	7.03621
9		—F	—CH_3	—H	—Cl	7.02687
10		—H	—CH_3	—H	—Cl	6.66756
11		—H	—CH_3	—H	—Cl	6.96658
12		—H	—CH_3	—H	—Cl	8.09691
13		—F	—CH_3	—H	—Cl	7.82391
14		—H	—CH_3	—H	—Cl	7.18046
15		—H	—CH_3	—H	—Cl	7.1549

Table 1 – *continued*

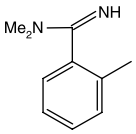
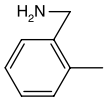
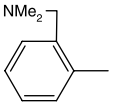
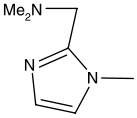
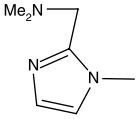
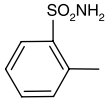
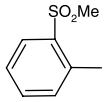
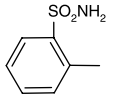
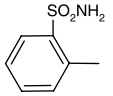
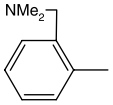
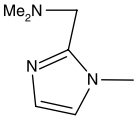
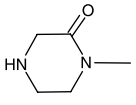
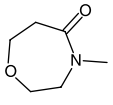
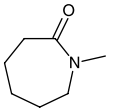
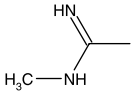
Number	R_1	R_2	R_3	R_4	R_5	PIC_{50}
16		—H	—CH_3	—H	—Cl	7.58503
17		—H	—CH_3	—H	—Cl	7.40894
18		—F	—CH_3	—H	—Cl	8.69897
19		—H	—CH_3	—H	—Cl	8.09691
20		—F	—CH_3	—H	—Cl	8.39794
21		—F	—CH_3	—F	—H	7.3279
22		—F	—CH_3	—F	—H	7.56864
23		—F	—CH_3	—SO_2Me	—H	7.65758
24		—F	—CH_3	—SO_2Me	—H	7.46852
25		—F	—CH_3	—SO_2Me	—H	6.95861
26		—F	—CH_3	—SO_2Me	—H	8.09691
27		—F	—CH_3	—SO_2Me	—H	7.38722
28		—F	—CH_3	—SO_2Me	?H	7.63827
29		—F	—CH_3	—SO_2Me	—H	7.3279
30		—F	—CH_3	—F	—H	6.9431

Table 1 – *continued*

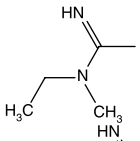
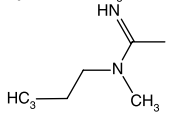
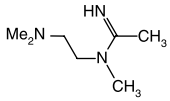
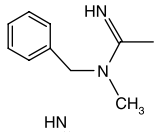
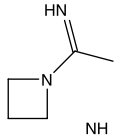
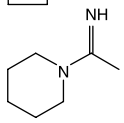
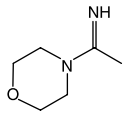
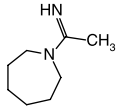
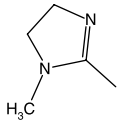
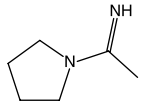
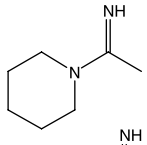
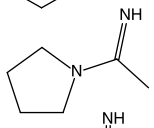
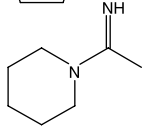
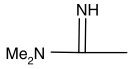
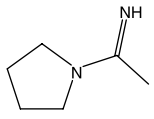
Number	*R_1*	*R_2*	*R_3*	*R_4*	*R_5*	*PIC_{50}*
31		—F	—CH_3	—F	—H	7.65758
32		—F	—CH_3	—F	—H	7.46852
33		—F	—CH_3	—F	—H	7.31876
34		—F	—CH_3	—F	—H	7.09691
35		—F	—CH_3	—F	—H	7.30103
36		—F	—CH_3	—F	—H	8.09691
37		—F	—CH_3	—F	—H	6.8729
38		—F	—CH_3	—F	—H	7.30103
39		—F	-CF_3	—F	—H	7.50864
40		—F	-CF_3	—H	—Cl	8.1549
41		—F	-CF_3	—H	—Cl	8.1549
42		—F	—CH_3	—H	—Cl	8.69897
43		—F	—CH_3	—SO_2Me	—H	7.55284
44		—F	-CF_3	—H	—Cl	9
45		—F	-CF_3	—H	—Cl	8.69897

Table 1 – *continued*

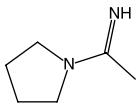
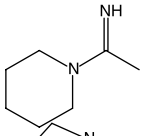
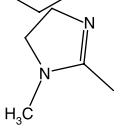
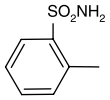
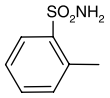
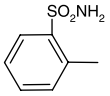
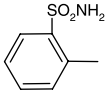
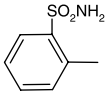
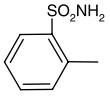
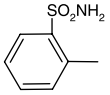
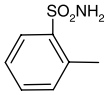
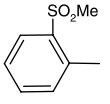
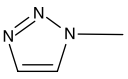
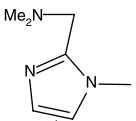
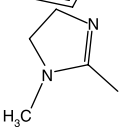
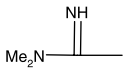
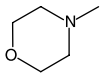
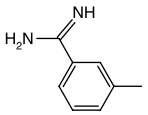
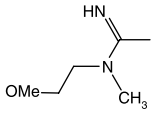
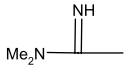
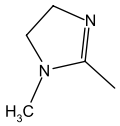
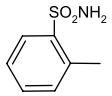
Number	R_1	R_2	R_3	R_4	R_5	PIC_{50}
46		—F	—CH_3	—SO_2Me	—H	7.82391
47		—F	—CH_3	—SO_2Me	—H	7.55284
48		—F	—CH_3	—SO_2Me	—H	7.85387
49		—H	—CH_3	—H	—H	6.98297
50		—H	—CH_3	-CH_2NH_2	—H	7.42022
51		—H	-CF_3	-$CONH_2$	—H	7.61979
52		—H	—CH_3	—SO_2NH_2	—H	7.61979
53		—H	—CH_3	—H	-Br	7.74473
54		—H	-CF_3	—H	—Cl	8.39794
55		—H	-CF_3	—F	—H	7.56864
56		—F	—CH_3	—H	—Cl	8.52288
57		—F	—CH_3	—H	—Cl	8.22185
58		—H	—CH_3	—H	—Cl	6.95078
59		—F	—CH_3	—F	—H	7.65758
60		—F	-CF_3	—H	—Cl	8.09691

Table 1 – *continued*

Number	*R_1*	*R_2*	*R_3*	*R_4*	*R_5*	*PIC_{50}*
61		—F	—CH_3	—SO_2Me	—H	7.92082
62		—H	—CH_3	—H	—Cl	7.05552
63		—H	—CH_3	—H	—Cl	7.12494
64		—F	—CH_3	—F	—H	6.95861
65		—F	—CH_3	—H	—Cl	8.30103
66		—F	—CH_3	—SO_2Me	—H	7.85387
67		—H	—CH_3	—$C(NH)NH_2$	—H	7.67778

2.4 Artificial neural networks

ANN is used to investigate QSARs extensively [14,15]. Most of the applications of neural networks in QSAR studies used fully connected three-layer, back-propagation neural networks (BPNN). The first layer is the input layer, and each of its neurons receives information from the exterior, corresponding to one of the independent variables used as inputs. The last layer is the output layer, and its neurons handle the output from the network. The layers of neurons between the input and output layers are called hidden layers. Each layer may make its independent computations and may pass the results yet to another layer [16]. With nonlinear transfer function in the hidden layer, it can provide excellent performances in many applications of fitting and reproducing almost any nonlinear models.

Table 2. The calculated descriptors used in this study.

Descriptor	*Brief Description*
ET	Total molecular energy
EB	Molecular binding energy
EI	Molecular isolated atomic energy
EE	Molecular electronic energy
Q_1	Net charge in position 1
Q_2	Net charge in position 2
Q_3	Net charge in position 3
Q_4	Net charge in position 4
Q_5	Net charge in position 5
EC	Molecular core–core interaction energy
ECI	CI energy
HF	Heat of molecular formation
DM	Total molecular dipole moment
SM	Molecular surface area
VM	Molecular volume
EH	Hydration energy
LOGP	*n*-Octanol/water partition
RM	Molecular molar refractivity
PM	Polarizability of molecule
MM	Mass of molecule
HOMO	Energy of the highest occupied molecular orbital
LUMO	Energy of the lowest unoccupied molecular orbital

The number of neurons in the hidden layer is an important factor determining the network's quality. That is, networks with few nodes may be insufficient to use all the information from the data (underfitting). In this case, we must increase neurons in the hidden layer. Too many nodes cause the network to memorize the data (overfitting), this often worsens the ability of generalization [17]. The MSE value was used to evaluate the quality of the networks. To evaluate the competence of a neural network, a validation process must be done. That is to say, a validation set must be used to check out the networks when they are training. If the difference between the MSE of the training set and the validation set is large, the model is overfitted. If the overfitting occurs, we must reduce the neurons in the hidden layer and re-train the networks until

the difference between the MSE of the training set and the validation set is small.

2.5 Genetic algorithm optimized ANN

GA is a kind of stochastic search algorithm inspired by the mechanics of natural evolution, including survival of the fittest, reproduction, cross-over, and mutation. GA is based on Darwinian survival of the fittest strategy, and work with a population of individuals, it uses selection and recombination operators to generate new sample points in a search space [18]. Definition of the search space and the search goal are problem dependent. For most QSAR models the search space is usually a space of solutions for a specified problem, and the goal is to identify the solution which performs the best according to an evaluation function, such as, for example, the MSE generated by training set or the testing set [19].

A simple GA usually consists of three processes selection, genetic operation and replacement. The population comprises a group of chromosomes that are the candidates for the solution. The fitness values of all chromosomes are evaluated using an objective function (performance criteria or a system's behavior) in a decoded form. A particular group of parents is selected from the population to generate offspring by the defined genetic operations of crossover and mutation. The fitness of all offspring is then evaluated using the same criterion and the chromosomes in the current population are then replaced by their offspring, based on a certain replacement strategy. Such a GA cycle is repeated until a desired termination criterion is reached. If all goes well throughout this process of simulated evolution, the best chromosome in the final population can become a highly evolved and more superior solution to the problem [20].

With this strategy, GA can then be used effectively in the evolution to find a near-optimal set of connection weights globally. The combination of GA and neural network (briefly described as GA-NN) for weight training consists of three major phases. First, binary strings form was used to represent connection weights. The second step is the evaluation on the fitness of these connection weights by constructing the corresponding neural network through decoding each genome and computing its fitness function and mean square error function. The third step is applying the evolutionary process such as selection, crossover, and mutation operations by a GA according to its fitness. The evolution stops when the fitness is greater than a predefined value [20].

3. Results and discussion

In this section, the prediction performances of PLS, ANN and GA-NN are evaluated. First, the datasets were normalized between the range of [−1,1]. After the networks have been trained, the outputs need to be transferred back to the same units that were used for the original outputs for comparison purpose. To ensure a fair comparison, the same training and test set are used for each of the models.

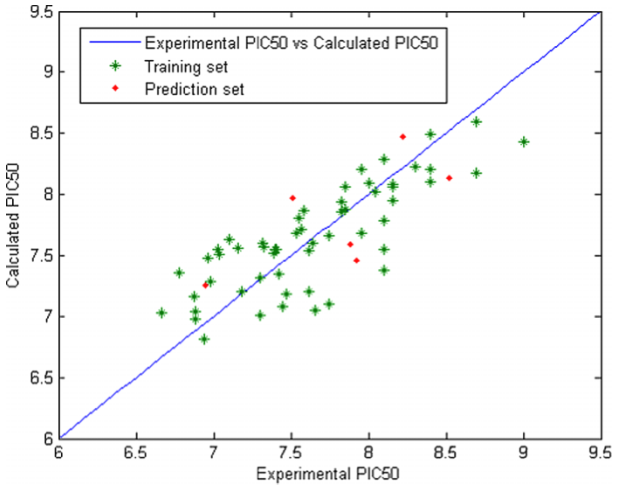


Figure 3. The scatterplot of observed vs. experimental PIC_{50} values of PLS model.

PLS analysis was performed by using 22 descriptors and 11 components. The training set was used to build up the PLS model, and the testing set was used to test the prediction ability of the model. As mentioned ahead, A leave-one-out cross validation procedure was used. From the results of the PLS regression, we got that the square of the correlation coefficient (R^2) and the RMSE of the training set was 0.6470 and 0.3222, respectively. The R^2 and the RMSE of the testing set was 0.4411 and 0.4588, respectively. The results are not satisfactory, it indicates that a linear model can not explain the relationship between the activities and the descriptors well. The scatterplot of observed vs. predicted PIC_{50} values is shown in figure 3.

In this paper, we built up a BPNN in MATLAB (version 7.0, MathWorks, Inc.) The training set was used to build the model and the testing set was used to test the prediction ability of the model. The validation set was used during the training process of ANN model to prevent neural network overfitting. In the program, we used principal component analysis function, because some descriptors have significant correlation, it can affect the model when the networks is learning.

In order to get the best model, several models with different number of hidden layers were built up. The models are summarized in table 3. The model with 6 hidden neurons which have the minimum RMSE of training set and testing set was selected as the best nonlinear model. The testing set was used to test the prediction ability of the model. From the results of the model, we got that the R^2 and the RMSE of the

Table 3. The result of different hidden neurons on the ANNs' performance.

Number	*RMSEtr*	*RMSEpre*
3	0.2374	0.3305
4	0.2268	0.3044
5	0.2164	0.2972
6	0.1788	0.2288
7	0.2277	0.2751
8	0.2253	0.3048
9	0.2460	0.3412

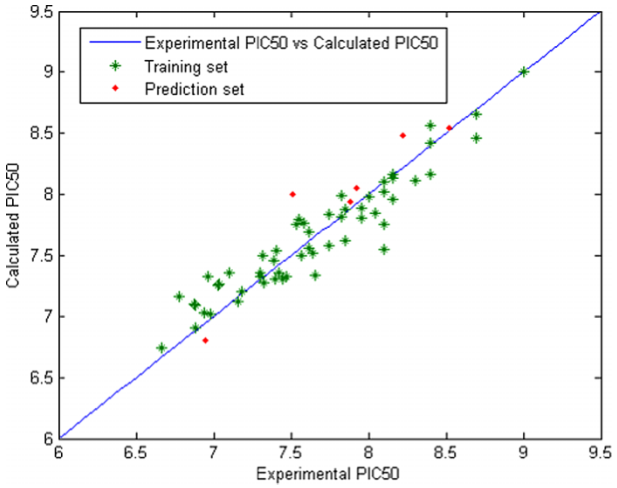


Figure 4. The scatterplot of calculated values vs. experimental values of PIC_{50} of the neural networks model.

training set was 0.8922 and 0.1788, respectively. The R^2 and the RMSE of the testing set was 0.8815 and 0.2288, respectively. The scatterplot of calculated vs. experimental values of PIC_{50} from the best nonlinear model are shown in figure 4.

The GA-NN program was carried out in MATLAB (version 7.0, MathWorks, Inc.) The BPNN were used to evaluate all chromosomes, that was, after the parameter values for each chromosome were translated into the network, it was trained on the training set, and the validation set was used to test if the network was overfitting or not. The training process of the network stops after a maximum of 1000 epochs or until there was no improvement of the RMSE for 200 epochs on validation set. The fitness of every chromosome was evaluated by measuring the RMSE. The test set was used to test the prediction ability of the model. From the results of the model, we got that the R^2 and the RMSE of the training set was 0.9235 and 0.1535, respectively. The R^2 and the RMSE of the testing set was 0.9056 and 0.1736, respectively. The scatterplot of

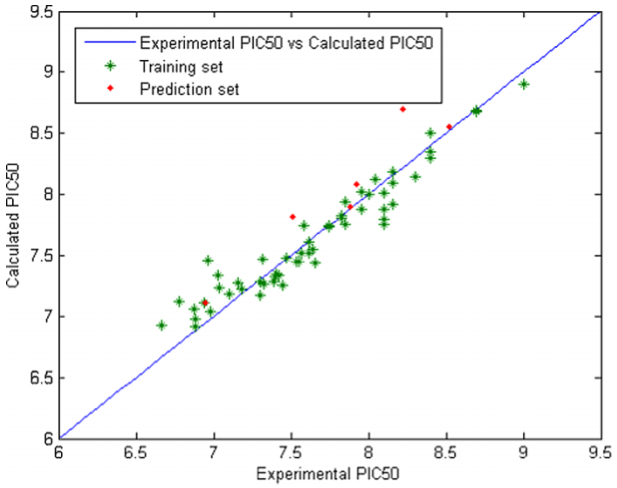


Figure 5. The scatterplot of calculated vs. experimental values of PIC_{50} of the GA-NN model.

Table 4. The results of PLS, ANN, GA-NN.

	Training set		*Testing set*	
Model	*RMSE*	R^2	*RMSE*	R^2
PLS	0.3222	0.6470	0.4588	0.4411
ANN	0.1788	0.8922	0.2288	0.8815
GA-ANN	0.1535	0.9235	0.1736	0.9056

calculated vs. experimental values of PIC_{50} from the model are shown in figure 5.

From the QSAR models of fXa inhibitors built up by PLS, ANN, GA-NN methods, we found that every model generated different results. The results of every model are given in table 4.

Taking account of the RMSE from table 4, the RMSE and R^2 for the training set is ordered by the following: GA-ANN gave the best results (0.1535 and 0.9235); ANN showed better results (0.1788 and 0.8922); PLS gave a poor result (0.3222 and 0.6470). For testing, the RMSE and R^2 for the training set is ordered by the following: GA-ANN gave the best results (0.1736 and 0.9056); ANN showed better results (0.2288 and 0.8815); PLS provided a poor result (0.4588 and 0.4411). The value for training set and testing set of PLS model are both greater than 0.3, but for ANN and GA-NN, which are nonlinear models, the value are both lower than 0.23. It indicates that for substituted 1-(2-naphthyl)-1H-pyrazole-5-carboxylamide fXa inhibitors, the nonlinear QSAR models show better performance than linear model.

The RMSE for training set and testing set of GA-NN model are both lower than ANN model. As discussed ahead, the GA optimized the ANN connection weights, and it generated a better result than simple ANN.

4. Conclusions

Thrombosis-related ischemic diseases are leading causes of death in the world. A particularly attractive new class of substituted 1-(2-naphthyl)-1H-pyrazole-5-carboxylamide fXa inhibitors appear to provide an enhanced pharmacodynamic action than conventional anticoagulation therapies.

In this paper, we built up the QSAR models of the substituted 1-(2-naphthyl)-1H-pyrazole-5-carboxylamide fXa inhibitors by means of three methods (PLS, ANN, and GA-NN). The GA-NN model shows the best result, the ANN model shows much better than the PLS model. The PLS model is a kind of linear model, it is suitable for modeling QSAR if the relationship between the descriptors and the experimental activities is linear. If the relationship is not linear, we would better use nonlinear methods to build up the QSAR model. The ANN and the GA-NN methods are useful methods to build up nonlinear models. Particularly, if we use the GA to optimize the connection weights of ANN, we can get a better result than simple ANN generally.

We built up three 2D QSAR models in this paper, and found the performance of nonlinear model is better than linear model. The GA-NN methods show better performance than ANN model. In our later study, we are going to investigate the 3D QSAR of these compounds, and compare the 2D and 3D QSAR methods. The aim of our research is to design new substituted 1-(2-naphthyl)-1H-pyrazole-5-carboxylamide fXa inhibitors.

Acknowledgements

This work was partially supported by West China Medical College.

References

[1] R.J. Leadley. Coagulation factor Xa inhibition: biological background and rationale. *Curr. Top. Med. Chem.*, **1**, 151 (2001).
[2] K. Sato, T. Kawasaki, Y. Taniuchi, F. Hirayama, H. Koshio, Y. Matsumoto. A novel factor Xa inhibitor: separation of its antithrombotic effects from its prolongation of bleeding time. *Eur. J. Pharmacol.*, **339**, 141 (1997).
[3] A.R. Katritzky, L.M. Pacureanu, S. Slavov, D.A. Dobchev, M. Karelson. QSAR study of antiplatelet agents. *Bioorg. Med. Chem.*, **14**, 7490 (2006).
[4] Z.J. Jia, Y. Wu, W. Huang, E. Goldman, P. Zhang, J. Woolfrey, P. Wong, B. Huang, U. Sinha, G. Park, A. Reed, R.M. Scarborougha, Y. Zhua. Design, synthesis and biological activity of novel non-amidine factor Xa inhibitors. Part 1: P[1] Structure–activity relationships of the substituted 1-(2-naphthyl)-1H-pyrazole-5-carboxylamides. *Bioorg. Med. Chem. Lett.*, **12**, 1651 (2002).
[5] W. Huang, P. Zhang, J.F. Zuckett, L. Wang, J. Woolfrey, Y. Song, Z.J. Jia, L.A. Clizbe, T. Su, K. Tran, B. Huang, P. Wong, U. Sinha, G. Park, A. Reed, J. Malinowski, S.J. Hollenbach, R.M. Scarborough, B.Y. Zhu. Design, synthesis and structure–activity relationships of benzoxazinone-based factor Xa inhibitors. *Bioorg. Med. Chem. Lett.*, **13**, 561 (2003).
[6] P. Jakobsen, B. Ritsmar Pedersen, E. Persson. Inhibitors of the tissue factor/factor VIIa-induced coagulation: synthesis and in vitro evaluation of novel specific 2-aryl substituted 4H-3,1-benzoxazin-4-ones. *Bioorg. Med. Chem.***8**, pp. 2095 [b] Costi, M. P. Second Joint Italian–Swiss Meeting on Medicinal Chemistry (ITCHMC2005) Modena, Italy, Sep (2005) (2000).
[7] D. Hadjipavlou-Litina. Activity studies of anticonvulsants. *Med. Res. Rev*, **18**, 91 (1998).
[8] D.E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, New York (1989).
[9] Z.J. Jia, Y. Wu, W. Huang, P. Zhang, L.A. Clizbe, E.A. Goldman, U. Sinha, A.E. Arfsten, S.T. Edwards, M. Alphonso, A. Hutchaleelaha, R.M. Scarborough, B.-Y. Zhu. 1-(2-Naphthyl)-1H-pyrazole-5-carboxylamides as potent factor Xa inhibitors. Part 2: a survey of P4 motifs. *Bioorg. Med. Chem. Lett.*, **14**, 1221 (2004).
[10] Z.J. Jia, Y. Wu, W. Huang, P. Zhang, Y. Song, J. Woolfrey, U. Sinha, A.E. Arfsten, S.T. Edwards, A. Hutchaleelaha, S.J. Hollennbach, J.L. Lambing, R.M. Scarborough, B.-Y. Zhu. 1-(2-Naphthyl)-1H-pyrazole-5-carboxylamides as potent factor Xa inhibitors. Part 3: design, synthesis and SAR of orally bioavailable benzamidine-P4 inhibitors. *Bioorg. Med. Chem. Lett.*, **14**, 1229 (2004).
[11] S. Wold, H. Martens, H. Wold. The multivariate calibration method in chemistry solved by the PLS method. In *Proceedings on the Conference on Matrix Pencils*, A. Ruhe, B. Kagstrm (Eds.), Lecture Notes in Mathematicspp. 286–293, Springer-Verlag, Heidelberg (1983).
[12] M.P. Freitas, J.A. Martins. Simple and highly predictive QSAR method: application to a series of (*S*)-*N*-[(1-ethyl-2-pyrrolidinyl)-methyl]-6-methoxybenzamides. *Talanta*, **56**, 182 (2005).
[13] K. Hasegawa, Y. Miyashita, S. Sasaki. Quantitative structure–activity relationship study of antiarrhythmic phenylpyridines using multivariate partial least squares modelling. *Chem. Int. Lab. Syst.*, **16**, 69 (1992).
[14] J.J. Huuskonen, D.J. Livingstone, I.V. Tetko. Neural network modeling for estimation of partition coefficient based on atom-type electrotopological state indices. *J. Chem. Inf. Comput. Sci.*, **40**, 947 (2000).
[15] A. Breindl, N. Beck, T. Clark, R.C. Glen. Prediction of the *n*-octanol/water partition coefficient, logP, using a combination of semiempirical MO-calculations and a neural network. *J. Mol. Model.*, **3**, 142 (1997).
[16] S.P. Niculescu. Artificial neural networks and genetic algorithms in QSAR. *J. Mol. Struct.*, **622**, 71 (2003).
[17] I. Kövesdi, M.F. Dominguez-Rodriguez, L. Ôrfi, G. Náray-Szabó, A. Varró, J.Gy. Papp, P. Mátyus. Application of neural networks in structure–activity relationships. *Med. Res. Rev.*, **19**, 249 (1999).
[18] S. Li, Y. Li, Y. Liu, Y. Xu. A GA-based NN approach for makespan estimation. *Appl. Math. Comput.*, **185**, 1003 (2007).
[19] D.F. Cook, C.T. Ragsdale, R.L. Major. Combining a neural network with a genetic algorithm for process parameter optimization. *Eng. Appl. Artif. Intel.*, **13**, 391 (2000).
[20] Z. Liu, A. Liu, C. Wang, Z. Niu. Evolving neural network using real coded genetic algorithm (GA) for multispectral image classfication. *Future Generat. Comput. Syst.*, **20**, 1119 (2004).