

## Research Article

# Prediction on the Inhibition Ratio of Pyrrolidine Derivatives on Matrix Metalloproteinase Based on Gene Expression Programming

Yuqin Li,<sup>1</sup> Guirong You,<sup>1</sup> Baoxiu Jia,<sup>1</sup> Hongzong Si,<sup>2</sup> and Xiaojun Yao<sup>3</sup>

<sup>1</sup> College of Pharmacy, Taishan Medical University, Taian, Shandong 271016, China

<sup>2</sup> Institute of Computer Science and Engineering Technology, Qingdao University, Qingdao 266071, China

<sup>3</sup> College of Chemistry and Chemical Engineering, Lanzhou University, Lanzhou 271000, China

Correspondence should be addressed to Yuqin Li; [liyqin@tsmc.edu.cn](mailto:liyqin@tsmc.edu.cn)

Received 27 February 2014; Accepted 29 April 2014; Published 22 May 2014

Academic Editor: Nick V. Grishin

Copyright © 2014 Yuqin Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Quantitative structure-activity relationships (QSAR) were developed to predict the inhibition ratio of pyrrolidine derivatives on matrix metalloproteinase via heuristic method (HM) and gene expression programming (GEP). The descriptors of 33 pyrrolidine derivatives were calculated by the software CODESSA, which can calculate quantum chemical, topological, geometrical, constitutional, and electrostatic descriptors. HM was also used for the preselection of 5 appropriate molecular descriptors. Linear and nonlinear QSAR models were developed based on the HM and GEP separately and two prediction models lead to a good correlation coefficient ( $R^2$ ) of 0.93 and 0.94. The two QSAR models are useful in predicting the inhibition ratio of pyrrolidine derivatives on matrix metalloproteinase during the discovery of new anticancer drugs and providing theory information for studying the new drugs.

## 1. Introduction

The tumor cell metastasis is a complex process that involves a series of processes such as the adhesion, enzymatic degradation, chemotaxis, and blood vessel hyperplasia in matrix [1]. Although there are many factors that affect the metastasis process of the malignant tumor cells, the interactive protein-degrading enzyme of the tumor cells and the surrounding microenvironment plays a key role in the deterioration of the tumor, which cannot be ignored [2]. Matrix metalloproteinases (MMPs) are one of them [3]. MMPs are a kind of endoenzyme depending on the zinc ion, playing an important role in the degradation and reconstruction of the extracellular matrix [4]. It turns out that MMPs play a crucial role in the tumor growth, invasion, metastasis, and angiogenesis in cancer tissue, in which gelatinase (MMP-2, MMP-9) is closely related to malignant tumors. Gelatinase (MMP-2, MMP-9) is an important target spot for antineoplastic drug research [5]. At present, it has been the hotspot of cancer drug research to develop and find the selective inhibitors of these target spots.

As a kind of alkaloid, with the derivative widely applied, pyrrolidine can be used as an important intermediate of fine chemicals and widely applied to the fields such as pharmaceutical [6, 7], food, pesticides, daily chemicals [8], paints, textiles, printing and dyeing, papermaking, photographic materials, and polymer materials. Recent studies have found that it has anticancer activity with its mechanism of action to inhibit the activity of MMP-2 and MMP-9 and thereby to inhibit the tumor growth, invasion, metastasis, and angiogenesis in cancer tissues.  $IC_{50}$  (the molar concentration of the compound leading to 50% enzyme inhibition) is often used to evaluate the effectiveness of the drug, as the action mechanism and therapeutic role of the drug after entering the body are closely related to its chemical structure and nature. However, these natures can be calculated or predicted by various methods. Quantitative structure-activity relationship (QSAR) and its variations have become a potentially effective way to predict the drug activity parameters [9–12]. The advantages of QSAR lie in that once the model is established, the nature of the compound can be predicted

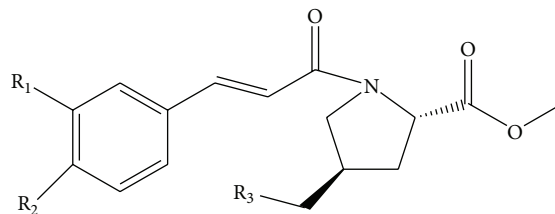


FIGURE 1: The common structure of compounds.

by the compound structure, and reasonable explanation can be made on the action mechanism of the drugs [13–15]. The method extends the range of rational drug screening and is helpful for finding new drugs according to the action mechanism [16–20].

Gene expression programming (GEP) [21] is a high efficient exploration algorithm based on the genetic evolution mechanism of natural population. Regarding the possible solutions in the problem domain as an individual or a chromosome of the group, coding the individual into the form of symbol string, carrying out repeated operation on the group based on the genetics (genetics, intersection, and heteromorphosis), evaluating the individuals according to the scheduled target fitness function, constantly obtaining better groups according to the evolution rule of “survival of the fittest,” and, meanwhile, searching the optimum individual with the searching approach in the overall situation to obtain the satisfied and optimal solutions, it has extremely strong generalization ability and has been used for the QSAR study of the drug [22–25].

This study adopts the heuristic algorithm (HM) and GEP to establish the QSAR model of pyrrolidine derivatives, gelatinase:  $IC_{50}$ , establish linear and nonlinear models, predict the  $IC_{50}$  of 33 pyrrolidine compounds, and also discuss the structural factors that affect the  $IC_{50}$ .

## 2. Data Set, Generation of Molecule Descriptors, and Methods

**2.1. Data Set.** The structures of the 33 pyrrolidine compounds (Figure 1) adopted and their corresponding  $IC_{50}$  values are from [26] and are listed in Table 1 with the logarithm collected. In the study of HM and GEP, the data set is randomly divided into two sets: the training set contains 21 compounds and is used to establish the models; and the test set contains 12 compounds used to evaluate the stability and predictive ability of the established models.

**2.2. Generation of Molecule Descriptors.** The two-dimensional structure of the molecules is drawn with the software ISISDRAW2.4. In the software Hyperchem7.0, all compounds shall be primarily optimized with the molecule mechanics method MM+, experiencing the geometry optimization with the semiempirical AM1 method on this basis to obtain the lowest energy conformation. The optimized molecule structure shall be calculated in the program MOPAC 7.0, with the resulting file of the MOPAC transferred into the program CODESSA to compute the five categories of descriptors,

namely, the structure, topology, geometry, electrostatic, and quantum chemical descriptors, with totally 496 descriptors obtained.

### 2.3. Methods

**2.3.1. HM Method.** The HM in the software CODESSA can realize the full search of a large number of molecule descriptors, so as to establish the optimum linear regression equation. The method firstly performs the colinearity control on the molecule descriptor with any two descriptors with the correlation coefficient higher than 0.8 and not being simultaneously contained in the same model, carries out rapid screening on the parameters with the heuristic method, and establishes the optimum model instead of examining all possible combinations of parameters. HM excludes some descriptors according to the following 4 rules: (1) parameters not common for each compound; (2) descriptors with relatively smaller value changes for all compounds; (3) parameters with the  $F$  test value less than 1.0 in an equation related to the parameter; and (4) descriptors with Student’s  $t$ -test value less than a defined value. The quality of the model shall be inspected by the correlation coefficient ( $R$ ), test value ( $F$ ), and the standard deviation ( $s$ ). The stability of the model shall be inspected by the leave-one-out (LOO) cross-validation correlation coefficient  $R_{CV}^2$ . In this study, the HM regression result is represented with the root mean square (RMS).

**2.3.2. GEP Method.** GEP is a new genetic algorithm invented by a Portuguese scientist in 1999 based on the genome (genome, GA) and phenotype (phenotype, GP). GEP mainly includes two aspects: chromosomes and expression trees (ETs). ET is mainly used to express the genetic coding information of the chromosome. In GEP, there are two languages used: the language of genes and ETs. The implementation techniques of GEP mainly include encoding scheme,  $K$  expressions, selection operator, mutation operator, insert string operation, gene inversion, restructuring operator, polygene chromosome and the contiguous function, the standard function set and users-defined functions based on the frequent function set, and fitness function selection (Table 2). There are three kinds of fitness functions for the classic GEP method, and this paper adopts the fitness function based on the absolute error:

$$f_i = \sum_{j=1}^n \left( R - \left| \frac{P_{(ij)} - T_j}{T_j} \cdot 100 \right| \right), \quad (1)$$

TABLE 1: The experimental and predicted log(IC<sub>50</sub>) and their residues of pyrrolidine derivatives to matrix metalloproteinases in training and test sets with HM and GEP.

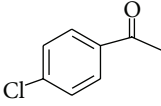
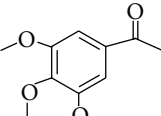
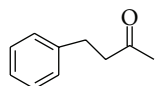
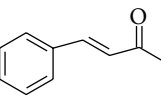
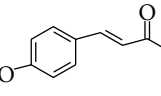
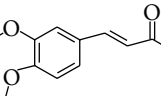
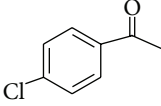
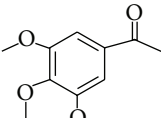
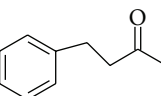
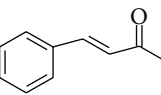
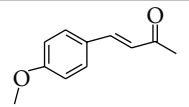
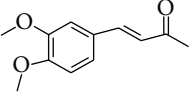
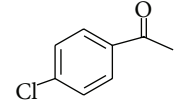
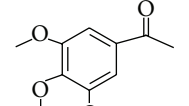
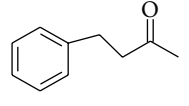
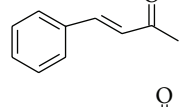
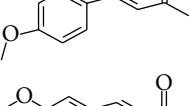
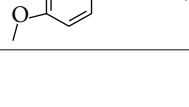
Number	Compound			Exp. log (IC <sub>50</sub> )	Calculated			
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>		HM	GEP		
					Pred. <sup>a</sup>	Residue <sup>b</sup>	Pred.	Residue
A0	H	H	H	1.0492	1.0885	-0.0393	0.9730	0.0762
A1	H	H	CH <sub>3</sub> CO	2.1086	1.9861	0.1225	2.3069	-0.1983
A2*	H	H	C <sub>2</sub> H <sub>5</sub> CO	1.9917	1.7100	0.2817	1.7100	0.2817
A3	H	H	C <sub>3</sub> H <sub>7</sub> CO	1.9325	1.9370	-0.0045	1.8882	0.0443
A4*	H	H	C <sub>6</sub> H <sub>5</sub> CO	1.7193	1.3271	0.3923	1.3271	0.3923
A5	H	H		1.5024	1.8876	-0.3852	1.7918	-0.2893
A6	H	H		2.4141	2.4248	-0.0107	2.3158	0.0983
A7*	H	H		1.6395	2.1844	-0.5450	2.1844	-0.5450
A8*	H	H		0.7160	0.6707	0.0453	0.6707	0.0453
A9	H	H		1.0899	1.0995	-0.0096	1.1843	-0.0944
A10	H	H		1.1173	1.3789	-0.2616	1.3999	-0.2826
B0	H	OMe	H	2.6433	2.6610	-0.0177	2.5506	0.0927
B1*	H	OMe	CH <sub>3</sub> CO	2.5002	2.8150	-0.3148	2.8150	-0.3148
B2	H	OMe	C <sub>2</sub> H <sub>5</sub> CO	2.4475	2.4473	0.0002	2.3449	0.1025
B3	H	OMe	C <sub>3</sub> H <sub>7</sub> CO	2.2900	2.1427	0.1473	2.4659	-0.1759
B4*	H	OMe	C <sub>6</sub> H <sub>5</sub> CO	2.0410	2.3434	-0.3024	2.3434	-0.3024
B5	H	OMe		1.6314	1.5272	0.1042	1.6977	-0.0662
B6	H	OMe		2.6990	2.7140	-0.0150	2.5052	0.1938
B7*	H	OMe		1.8657	2.1945	-0.3288	2.1945	-0.3288
B8	H	OMe		1.5922	1.4546	0.1376	1.5196	0.0726

TABLE I: Continued.

Number	Compound			Exp. log (IC <sub>50</sub> )	Calculated			
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>		HM Pred. <sup>a</sup>	Residue <sup>b</sup>	GEP Pred.	Residue
B9*	H	OMe		0.8921	0.7345	0.1576	0.7345	0.1576
B10	H	OMe		2.0839	1.7275	0.3564	1.6617	0.4222
C0*	OMe	OMe	H	1.9325	1.5873	0.3452	1.5873	0.3452
C1	OMe	OMe	CH <sub>3</sub> CO	2.5054	2.5197	-0.0143	2.3957	0.1098
C2	OMe	OMe	C <sub>2</sub> H <sub>5</sub> CO	2.4675	2.3163	0.1512	2.5349	-0.0674
C3*	OMe	OMe	C <sub>3</sub> H <sub>7</sub> CO	2.3446	2.7735	-0.4289	2.7735	-0.4289
C4	OMe	OMe	C <sub>6</sub> H <sub>5</sub> CO	2.3036	2.5582	-0.2546	2.4040	-0.1004
C5	OMe	OMe		2.0484	1.9164	0.1320	1.7615	0.2870
C6*	OMe	OMe		2.6990	2.4750	0.2240	2.4750	0.2240
C7	OMe	OMe		2.2261	2.3091	-0.0830	2.2892	-0.0631
C8	OMe	OMe		1.9370	1.6166	0.3204	1.7618	0.1752
C9*	OMe	OMe		1.4579	1.8722	-0.4143	1.8722	-0.4143
C10	OMe	OMe		0.9868	1.3632	-0.3764	1.2809	-0.2941

\*The compounds of the test set.

<sup>a</sup>The predicted log (IC<sub>50</sub>).

<sup>b</sup>Residue = log (Exp.) - log (Pred.).

where  $R$  is the selection range,  $P_{(ij)}$  is the predicted value by the individual program  $i$  for fitness case  $j$  (out of  $n$  fitness cases), and  $T_j$  is the target value for fitness case  $j$ .

### 3. Results and Discussions

**3.1. Calculation Results of HM.** All 33 compounds obtain 496 descriptors in total through the computing of the software CODESSA with all computed descriptors to establish the linear model for predicting log (IC<sub>50</sub>). To determine the appropriate number of descriptors, this research studies different sets of the descriptors. When there is no significant improvement in the statistical performance of the model

to add another descriptor, it means that the descriptor number is proper. The  $R^2$  increase of less than 0.02 or  $R_{CV}^2$  decrease shall be selected as the limit standard to avoid the "over parameterization" of the model. In this study, the five descriptors closely related to the inhibition rate are finally selected (Table 3). The correlation matrix of five descriptors is showed in Table 4. Seen from Table 4, the correlation coefficients between each of the two descriptors are less than 0.80, which means that they are interactively independent [27].

Figure 2 shows the correlation diagram of the predicted and experimental values of multiple linear regression models, which includes a total of 33 compounds of the training and test sets. The predicted log (IC<sub>50</sub>) of these compounds is also

TABLE 2: All the parameters and selection of GEP.

Parameters	Selection
Division	/
Addition	+
Square Root	Sqrt
Sine	Sin
Tangent	Tan
Multiplication	*
Subtraction	-
Power	Pow
Natural logarithm	Ln
10^X	Pow10
Chromosomes	100
Genes	5
Head size	8
Gene size	26
Linking function	Addition
Generations without change	200
Number of tries	3
Max. complexity	5
Error type	MSE
Precision	—
Selection range	—
0/1 rounding threshold	—
Mutation rate	0.044
Inversion rate	0.1
IS transposition rate	0.1
RIS transposition rate	0.1
One-point recombination rate	0.3
Two-point recombination rate	0.3
Gene recombination rate	0.1
Gene transposition rate	0.1
Constants per gene	10
Data type	Floating-point
Lower bound	-10
Upper bound	10
RNC mutation	0.01
Dc mutation	0.044
Dc inversion	0.1
Dc IS transposition	0.1

shown in Table 1. Finally, the linear QSAR model by the HM is gained as

$$\begin{aligned} \log(IC_{50}) = & -1.9501 \times 10^2 + 2.4570 \text{ LUMO} \\ & - 3.6715 \text{ MRECO} - 2.0681 \times 10^{-1} \text{ KSIND} \\ & - 7.0757 \text{ ZX} + 8.4804 \times 10^{-1} \text{ MASEOAT.} \end{aligned} \quad (2)$$

Train set:  $R^2 = 0.93$ ,  $R_{CV}^2 = 0.87$ ,  $F = 20.60$ , and  $s = 0.23$ .

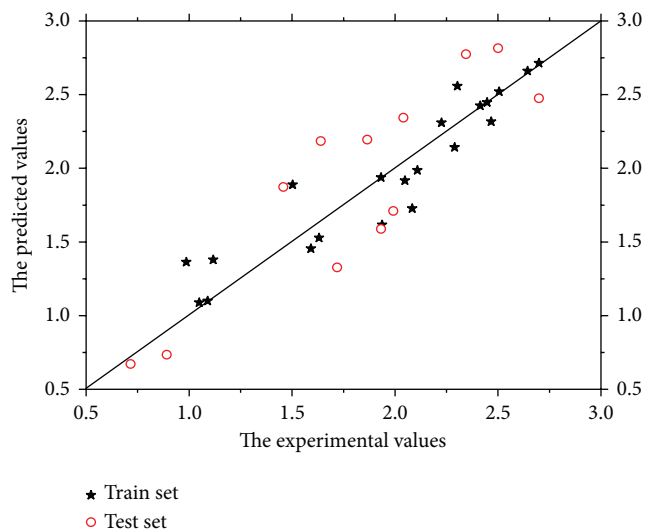


FIGURE 2: Plot of predicted log (IC<sub>50</sub>) versus experimental values for the training and test sets by HM.

Test set:  $R^2 = 0.85$ ,  $R_{CV}^2 = 0.50$ ,  $F = 21.13$ , and  $s = 0.36$ .

3.2. Calculation Results of GEP. After the establishment of the linear model, the same descriptors, as the variables of GEP, establish the nonlinear model. In order to obtain satisfactory results, the parameters affecting the GEP are optimized. Automatic problem solver (APS), the software package used by GEP, is easy to control, and therefore, the evolutionary model can be tested by the test set. In the course of evolution, good selection has been made for the functions with 7 functions selected, namely, subtract, multiply, divide, index, sin, and tan and the fitting function is MSE. Through fitting, the five descriptors selected establish the best QSAR model with the prediction values and residua listed in Table 1 and Figures 3 and 4. The nonlinear QSAR model by the GEP is gained as follows:

```
double dblTemp = 0.0,
dblTemp = sin (tan((tan (d[1])/sin (d[4])))),
dblTemp += sin (sin(((tan (d[1])/d[0])-d[3]))),
dblTemp += d[0],
dblTemp += pow (d[4],(pow (d[4],d[0])/d[2])),
dblTemp += sin (sqrt((d[2]-tan (sin(tan((d[2]*-
7.653931))))))),
```

where d[0], d(1), d(2), d(3), and d(4) represent LUMO, MRECO, KSIND, ZX, and MASEOAT, respectively. The statistical results of the established models are

Training set:  $R^2 = 0.94$ ,  $s = 0.12$ ;  
 Test set:  $R^2 = 0.81$ ,  $s = 3.95$ .

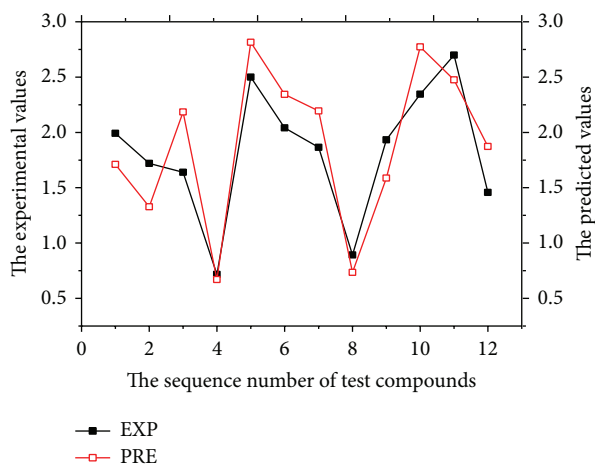
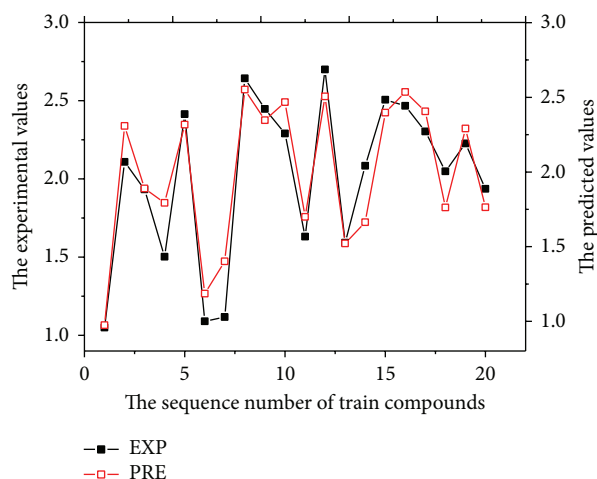
3.3. Discussions on Relevant Descriptor in the Model. By interpreting the model descriptors, the structural features

TABLE 3: Descriptors and their physical-chemical meanings, coefficient, error, and Student's  $t$ -test in HM.

Number	Descriptor	Physical-chemical meanings	Coefficient	Error	$t$ -test
0		Intercept	$-1.9501e + 02$	$1.2612e + 02$	$-1.5463$
1	LUMO	LUMO energy	$2.4570e + 00$	$5.0431e - 01$	$4.8720$
2	MRECO	Min resonance energy for a C–O bond	$-3.6715e + 00$	$6.7200e - 01$	$-5.4635$
3	KSIND	Kier shape index (order 3)	$-2.0681e - 01$	$7.7119e - 02$	$-2.6816$
4	ZX	ZX Shadow/ZX Rectangle	$-7.0757e + 00$	$2.1621e + 00$	$-3.2726$
5	MASEOAT	Min atomic state energy for a O atom	$8.4808e - 01$	$4.3585e - 01$	$1.9458$

TABLE 4: Correlation matrix of the 5 descriptors.

Descriptor	LUMO	MRECO	KSIND	ZX	MASEOAT
LUMO	1.0000				
MRECO	0.1497	1.0000			
KSIND	-0.5319	-0.4830	1.0000		
ZX	-0.0117	0.3261	0.1729	1.0000	
MASEOAT	0.1171	0.3261	-0.5478	0.6954	1.0000

FIGURE 3: Plot of predicted  $\log(IC_{50})$  versus experimental values for the training sets by GEP.FIGURE 4: Plot of predicted  $\log(IC_{50})$  versus experimental values for the test sets by GEP.

affecting the  $\log(IC_{50})$  values of these compounds may be identified. In the five parameters of the model selected, LUMO, MRECO, and MASEOAT are quantum chemistry descriptors; KSIND is a topological descriptor; and ZX is a geometric descriptor. The marshalling sequence of the descriptors in the equation shows that the contribution of the descriptor to  $\log(IC_{50})$  of the compound is in the order of LUMO > MRECO > KSIND > ZX > MASEOAT.

LUMO reflects the electron affinity of the molecule [28], with the coefficient positive in the model. When the target is fixed, the electrophilicity of the molecules is stronger, and the  $\log(IC_{50})$  value is greater. When  $R_3$  side chain is the aliphatic chain, the longer the chain, the greater the LUMO value, and the compound inhibition of enzyme activity of MMP-2 and MMP-9 will be increasing; the aromatics substituent is obviously stronger than the aliphatic substituent in side chain activity, which may be resulting from the large conjugation

system of the aromatic ring, increasing the LUMO value with stronger inhibition rate on the gelatinase activity. Generally, the substituent compound with branched chains is greater than that with a ring substituent, which means that the carbonyl reaction activity with open loop structure is stronger.

MRECO represents the minimum resonance energy of the C–O bond [29]. With the increase of the substituent, the three series of A, B, and C compounds keep an overall downward trend. The smaller the value, the lower the minimum resonance energy of the C–O bond, and the molecule is in a relatively stable state, highly reactive, and easy for the target combination. As its coefficient in the model is negative, with the decreasing of the MRECO, the value of  $\log(IC_{50})$  is gradually increased.

KSIND represents the three connectivity indexes of the molecule [30], represents the molecule size, shape, and degree of branching, and reflects the dispersion force between the

molecule volume and the molecules to a certain extent. The larger the molecule volume, the greater the molecule dispersion force. Table 2 shows that the KSIND value increases along with the increase of the atom number and structure of the substituent, and, therefore, the steric hindrance and dispersion force of the molecule also increase. The introduction of the group with large volume and strong rigidity is against the activity and the combination with the target decreases accordingly, leading to the  $\log(\text{IC}_{50})$  value decrease, which is in line with the negative coefficient in the model.

ZX represents the relative area of the projection part on the ZX plane of the molecule van der Waals [31], with Z and X representing the maximum and minimum inertial axes of the molecule, respectively. The appearing of the model descriptor means that the size of the molecule has great impacts on the  $\log(\text{IC}_{50})$  value of the drug, and the van der Waals force is an important part of the interaction energy between the subjects and objects. With negative coefficient in the model, the absolute value is relatively large, and, therefore, its increase results in the decrease of the  $\log(\text{IC}_{50})$  value of the drug. However, the compounds with structures similar to butterfly have higher flexibility and high activity.

MASEOAT [32] represents the minimum atomic state energy of the O atoms in the molecule and is related to the location of the oxygen atoms in the molecule, the molecule structure, and the steric hindrance. The lower the energy states of the oxygen atom, the higher its reactivity, and the easier the target molecule interactions. The description shows that the oxygen atoms in the molecule are related to the biological activity. In the model, the coefficient is positive, indicating that the energy state of the oxygen atom is positively correlated to the  $\log(\text{IC}_{50})$  value.

In summary, by comparing the data of in vitro inhibitory activities of the three series of A, B, and C, it can be seen that as A, B, and C molecule increases, the activity tends to decrease, suggesting that the smaller the side chain molecule of the  $R_1$  is, the more active the molecule is. The series of pyrrolidine compounds have good gelatinase inhibiting activity, and it is found that within a certain range, the larger the side chain of pyrrolidine ring C4, the better the flexibility, and the higher the activity; the activity of aromatic ring substituent is obviously higher than that of the aliphatic hydrocarbon substituent; and the compound with butterfly structure has higher activity.

## 4. Conclusions

This study proposes a method to predict the activity inhibition rate of pyrrolidine derivatives on gelatinase (MMP-2, MMP-9) based on HM and GEP. By calculating the molecule structure descriptors and establishing linear and nonlinear QSAR models by HM and GEP, the prediction results are satisfactory. Comparing the results of the two methods, we can see that both the linear HM method and nonlinear GEP method have strong predictive ability and better model stability in the activity inhibition rate of pyrrolidine derivatives on gelatinase (MMP-2, MMP-9), providing a theoretical basis for the in vitro screening of antitumor pyrrolidine derivatives.

## Abbreviations

HM:	Heuristic method
GEP:	Gene expression programming
Ets:	Expression trees
MMPs:	Matrix metalloproteinases
QSAR:	Quantitative structure-property relationships
Exp:	The experimental $\log(\text{IC}_{50})$
Pred:	The predicted $\log(\text{IC}_{50})$ .

## Conflict of Interests

The authors declare that they have no conflict of interests.

## Authors' Contribution

Yuqin Li conceived and designed the study, carried out data analysis, interpreted the entire results, and drafted the paper. Guirong You carried out data analysis of HM. Baoxiu Jia helped to draft the paper. Hongzong Si carried out data analysis of GEP. Xiaojun Yao participated in the design of the study and interpreted the results. All authors read and approved the final paper.

## Acknowledgment

This work was supported by the National Natural Scientific Foundation of Shandong Province (no. ZR2011BL007).

## References

- [1] I. Malanchi, "Tumour cells coerce host tissue to cancer spread," *BoneKey Reports*, vol. 2, p. 371, 2013.
- [2] C. Coghlin and G. I. Murray, "Current and emerging concepts in tumour metastasis," *Journal of Pathology*, vol. 222, no. 1, pp. 1-15, 2010.
- [3] E. Hadler-Olsen, J. O. Winberg, and L. Uhlin-Hansen, "Matrix metalloproteinases in cancer: their value as diagnostic and prognostic markers and therapeutic targets," *Tumor Biology*, vol. 334, no. 4, pp. 2041-2051, 2013.
- [4] X. Li, L. Qu, Y. Zhong, Y. Zhao, H. Chen, and L. Daru, "Association between promoters polymorphisms of matrix metalloproteinases and risk of digestive cancers: a meta-analysis," *Journal of Cancer Research and Clinical Oncology*, vol. 139, no. 9, pp. 1433-1447, 2013.
- [5] M. Verslegers, K. Lemmens, I. van Hove, and L. Moons, "Matrix metalloproteinase-2 and -9 as promising benefactors in development, plasticity and repair of the nervous system," *Progress in Neurobiology*, vol. 105, pp. 60-78, 2013.
- [6] X. Li and J. Li, "Recent advances in the development of MMPis and APNis based on the pyrrolidine platforms," *Mini-Reviews in Medicinal Chemistry*, vol. 10, no. 9, pp. 794-805, 2010.
- [7] X.-C. Cheng, Q. Wang, H. Fang, and W.-F. Xu, "Advances in matrix metalloproteinase inhibitors based on pyrrolidine scaffold," *Current Medicinal Chemistry*, vol. 15, no. 4, pp. 374-385, 2008.
- [8] S. Yu, J. Saenz, and J. K. Srirangam, "Facile synthesis of N-aryl pyrroles via Cu(II)-mediated cross coupling of electron deficient pyrroles and arylboronic acids," *Journal of Organic Chemistry*, vol. 67, no. 5, pp. 1699-1702, 2002.

- [9] R. Guha, "On exploring structure-activity relationships," *Methods in Molecular Biology*, vol. 993, pp. 81–93, 2013.
- [10] H. González-Díaz, S. Arrasate, N. Sotomayor et al., "MIANN models in medicinal, physical and organic chemistry," *Current Topics in Medicinal Chemistry*, vol. 13, no. 5, pp. 619–641, 2013.
- [11] M. Wesolowski and B. Suchacz, "Artificial neural networks: theoretical background and pharmaceutical applications: a review," *Journal of AOAC International*, vol. 95, no. 3, pp. 652–668, 2012.
- [12] J. Gálvez, M. Gálvez-Llompart, and R. García-Domenech, "Introduction to molecular topology: basic concepts and application to drug design," *Current Computer—Aided Drug Design*, vol. 8, no. 3, pp. 196–223, 2012.
- [13] P. Jasinski, P. Zwolak, R. Isaksson Vogel et al., "MT103 inhibits tumor growth with minimal toxicity in murine model of lung carcinoma via induction of apoptosis," *Investigational New Drugs*, vol. 29, no. 5, pp. 846–852, 2011.
- [14] D. V. Singh, S. Agarwal, R. K. Kesharwani, and K. Misra, "3D QSAR and pharmacophore study of curcuminoids and curcumin analogs: interaction with thioredoxin reductase," *Interdisciplinary Sciences*, vol. 5, no. 4, pp. 286–295, 2013.
- [15] Z. Yan, L. Zhang, H. Fu, Z. Wang, and J. J. Lin, "Design of the influenza virus inhibitors targeting the PA endonuclease using 3D-QSAR modeling, side-chain hopping, and docking," *Bioorganic & Medicinal Chemistry Letters*, vol. 24, no. 2, pp. 539–547, 2014.
- [16] G. Subramanian and S. N. Rao, "Comprehending renin inhibitor's binding affinity using structure-based approaches," *Bioorganic & Medicinal Chemistry Letters*, vol. 23, no. 24, pp. 6667–6672, 2013.
- [17] A. Manvar, V. Khedkar, J. Patel et al., "Synthesis and binary QSAR study of antitubercular quinolyldrazides," *Bioorganic & Medicinal Chemistry Letters*, vol. 23, no. 17, pp. 4896–4902, 2013.
- [18] S. Marchetti, D. Pluim, M. V. Eijndhoven et al., "Effect of the drug transporters ABCG2, Abcg2, ABCB1 and ABCC2 on the disposition, brain accumulation and myelotoxicity of the aurora kinase B inhibitor barasertib and its more active form barasertib-hydroxy-QPA," *Investigational New Drugs*, vol. 31, no. 5, pp. 1125–1135, 2013.
- [19] C. Ventura, D. A. Latino, and F. Martins, "Comparison of multiple linear regressions and neural networks based QSAR models for the design of new antitubercular compounds," *European Journal of Medicinal Chemistry*, vol. 70, pp. 831–845, 2013.
- [20] A. Worachartcheewan, C. Nantasenamat, W. Owasirikul et al., "Insights into antioxidant activity of 1-adamantylthiopyridine analogs using multiple linear regression," *European Journal of Medicinal Chemistry*, vol. 73, pp. 258–264, 2013.
- [21] P. Liu and W. Long, "Current mathematical methods used in QSAR/QSPR studies," *International Journal of Molecular Sciences*, vol. 10, no. 5, pp. 1978–1998, 2009.
- [22] H. Si, J. Zhao, L. Cui et al., "Study of human dopamine sulfo-transferases based on gene expression programming," *Chemical Biology and Drug Design*, vol. 78, no. 3, pp. 370–377, 2011.
- [23] W. Shi, X. Zhang, and Q. Shen, "Quantitative structure-activity relationships studies of CCR5 inhibitors and toxicity of aromatic compounds using gene expression programming," *European Journal of Medicinal Chemistry*, vol. 45, no. 1, pp. 49–54, 2010.
- [24] Y.-Q. Li, H.-Z. Si, Y.-L. Xiao et al., "Quantitative structure activity relationship models based on heuristic method and gene expression programming for the prediction of the  $pK_a$  values of sulfa drugs," *Chin Acta Pharm Sinica*, vol. 44, no. 5, pp. 486–490, 2009.
- [25] H. Si, N. Lian, S. Yuan et al., "Predicting the activity of drugs for a group of imidazopyridine anticoccidial compounds," *European Journal of Medicinal Chemistry*, vol. 44, no. 10, pp. 4044–4050, 2009.
- [26] L. Zhang, W.-F. Xu, and J. Zhang, "Protein catabolic enzymes inhibitory activities of pyrrolidine derivatives," *Chinese Pharmaceutical Journal*, vol. 43, no. 6, pp. 472–474, 2008.
- [27] D. F. Qiu, Q. Zhao, K. C. Liu, Y. C. Guo, and Y. Q. Feng, "Synthesis, crystal structure and biological activity of a planar copper complex derived from s-benzylthiocarbamate," *Chinese Journal of Structural Chemistry*, vol. 29, no. 10, pp. 1513–1518, 2010.
- [28] K. Tuppurainen, "Frontier orbital energies, hydrophobicity and steric factors as physical QSAR descriptors of molecular mutagenicity. A review with a case study: MX compounds," *Chemosphere*, vol. 38, no. 13, pp. 3015–3030, 1999.
- [29] A. G. Mercader, P. R. Duchowicz, F. M. Fernández, and E. A. Castro, "Modified and enhanced replacement method for the selection of molecular descriptors in QSAR and QSPR theories," *Chemometrics and Intelligent Laboratory Systems*, vol. 92, no. 2, pp. 138–145, 2008.
- [30] L. C. Porto, É. S. Souza, B. da Silva Junkes, R. A. Yunes, and V. E. F. Heinzen, "Semi-empirical topological index: development of QSPR/QSRR and optimization for alkylbenzenes," *Talanta*, vol. 76, no. 2, pp. 407–412, 2008.
- [31] P. Silakari, S. D. Shrivastava, G. Silakari et al., "QSAR analysis of 1,3-diaryl-4,5,6,7-tetrahydro-2H-isoindole derivatives as selective COX-2 inhibitors," *European Journal of Medicinal Chemistry*, vol. 43, no. 7, pp. 1559–1569, 2008.
- [32] H. Z. Si, T. Wang, K. J. Zhang, Z. D. Hu, and B. T. Fan, "QSAR study of 1,4-dihydropyridine calcium channel antagonists based on gene expression programming," *Bioorganic and Medicinal Chemistry*, vol. 14, no. 14, pp. 4834–4841, 2006.