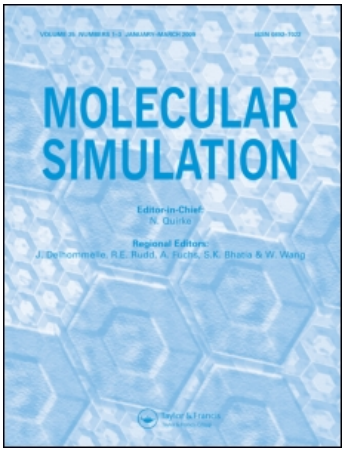


Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

QSAR model of triterpene derivatives as potent anti-HIV agents

Qing Su[a]; Xiaohong Xu[a]; Lu Zhou[b]
[a] Department of Pharmacy, Chengdu Medical College, Chengdu, People's Republic of China [b] Department of Chemical Engineering, Sichuan University, Chengdu, People's Republic of China

To cite this Article Su, Qing , Xu, Xiaohong and Zhou, Lu(2008) 'QSAR model of triterpene derivatives as potent anti-HIV agents', Molecular Simulation, 34: 7, 651 — 659
To link to this Article: DOI: 10.1080/08927020802036054
URL: http://dx.doi.org/10.1080/08927020802036054

Molecular Simulation
Vol. 34, No. 7, June 2008, 651–659

Taylor & Francis
Taylor & Francis Group

QSAR model of triterpene derivatives as potent anti-HIV agents

Qing Su[a], Xiaohong Xu[a]* and Lu Zhou[b]

[a]*Department of Pharmacy, Chengdu Medical College, Chengdu, People's Republic of China;* [b]*Department of Chemical Engineering, Sichuan University, Chengdu, People's Republic of China*

(Received 8 June 2007; final version received 5 March 2008)

Triterpenes derivatives are a promising class of anti-human immunodeficiency virus (anti-HIV) agents. A series of triterpene derivatives have been studied for quantitative structure–activity relationship with multiple linear regression (MLR) and artificial neural networks (ANN). The linear model with MLR performed poorly while the nonlinear model with ANN preformed well. For the ANN model with architecture of 5-6-1, the root mean square error for the training set, validation set and the prediction set are 0.2019, 0.2214 and 0.2883, respectively, and the r for the training set, validation set and the prediction set are 0.8895, 0.8843 and 0.8819, respectively. In this study, different methods were used to select the most relevant descriptors for MLR and ANN. Finally, five descriptors, EB, LOGP, HOMO, HE and ω, were selected for both MLR and ANN. The result indicates these descriptors are playing an important role on the anti-HIV activity of triterpene derivatives.

Keywords: triterpene derivatives; anti-HIV; artificial neural networks; QSAR

1. Introduction

Acquired immunodeficiency syndrome (AIDS), caused by human immunodeficiency virus (HIV) infection, has become an exceptional crisis because of both its emergent and long-term development. Current anti-AIDS drugs include inhibitors of reverse transcriptase, protease, and fusion. However, these existing drugs encounter problems such as the emergence of drug resistant viruses and toxicities caused by long-term drug usage [1,2]. Therefore, compounds possessing potent anti-HIV activity with novel modes of action are urgently needed to add to the existing anti-HIV therapies.

Triterpenes, including betulinic acid, oleanolic acid and ursolic acid, etc. (Figure 1), constitute a promising class of anti-HIV agents. Betulinic acid, isolated from the leaves of *Syzygium claviflorum* (Myrtaceae) etc., was identified as an anti-HIV agent in acutely infected H9 lymphocyte cells with an EC_{50} value of 1.4 μM and therapeutic index (TI) of 9.3 [3]. Subsequent modification of betulinic acid yielded analogues with greatly enhanced activity [4–9]. Betulinic acid derivative's modes of action have been associated with several steps in the virus life cycle, including fusion [10,11], reverse transcription [12,13] and maturation [14,15], depending primarily on the side chain structures of the compounds. The related oleanolic acid was isolated from *Rosa woodsii*, etc. displayed weak anti-HIV activity with EC_{50} values of 3.7 μM and TI values of 12.8. Ursolic acid, with an EC_{50} value of 4.4 μM, isolated from *Prosopis glandulosa* etc., showed a level of anti-HIV activity similar to that shown by oleanolic acid. As with betulinic acid, modification of oleanolic acid and ursolic acid also gave potent anti-HIV active derivatives [16].

Based on these previous findings, in this study, we collected 60 derivatives of betulinic acid, oleanolic acid and ursolic acid with their experimental IC_{50} value to build the quantitative structure–activity relationship (QSAR) models by multiple linear regression (MLR) and artificial neural networks (ANN); that is to model experimentally determined IC_{50} value from computationally-derived molecular descriptors of triterpene derivatives, to try to analyse the influence of descriptors and to offer guidance for the design of this series of compounds.

2. Material and methods

2.1 Data set

The biological data for 60 triterpene derivatives used in this study were collected from several papers (numbers 1–5 and 28–29 were obtained from [17], numbers 6–27 and 30–32 were obtained from [18], numbers 33–52 were obtained from [19], numbers 53–60 were obtained from [20]). The biological activity data IC_{50} (the concentration which inhibits H9 lymphocytes cell growth by 50%) of all the derivatives were tested with the same procedure, which can reduce the experimental error for the model. The IC_{50} expressed in μM (μMolar per

*Corresponding author. Email: xiaohong65523@yahoo.com.cn

ISSN 0892-7022 print/ISSN 1029-0435 online

DOI: 10.1080/08927020802036054
http://www.informaworld.com

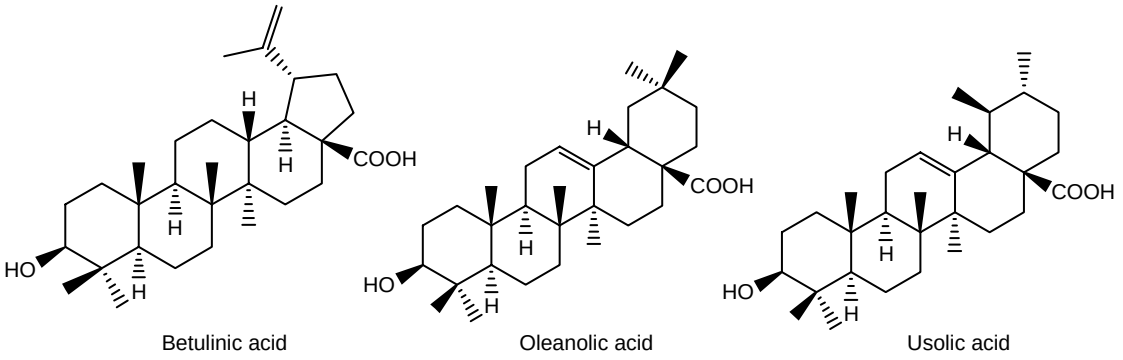


Figure 1. Chemical structures of betulinic acid, oleanolic acid and usolic acid.

milliliter) was converted to logarithm log ($1/IC_{50}$) (briefly described as PIC_{50}). When IC_{50} is given in an interval, the minimum value was used. The PIC_{50} was also called the dependent variable later. The triterpene skeletons of compounds (1)–(58) and the chemical structure of compounds (59)–(60) are shown in Figure 2, the substitutions for different triterpene skeleton are listed in Tables 1–7, and the experimental IC_{50} and PIC_{50} of 60 triterpene derivatives are listed in Table 8.

Descriptors derived from quantum chemical computation can clearly describe molecular structure and electronic properties: therefore, quantum chemical descriptors included in general QSAR models are popular for the development of reliable QSAR models. Descriptors used in this study were calculated with HyperChem 8, HyperCube, Inc. Firstly, MM^+ molecular mechanics force field was run to get close to the optimised geometry. Secondly, the conformation obtained from molecular mechanics was

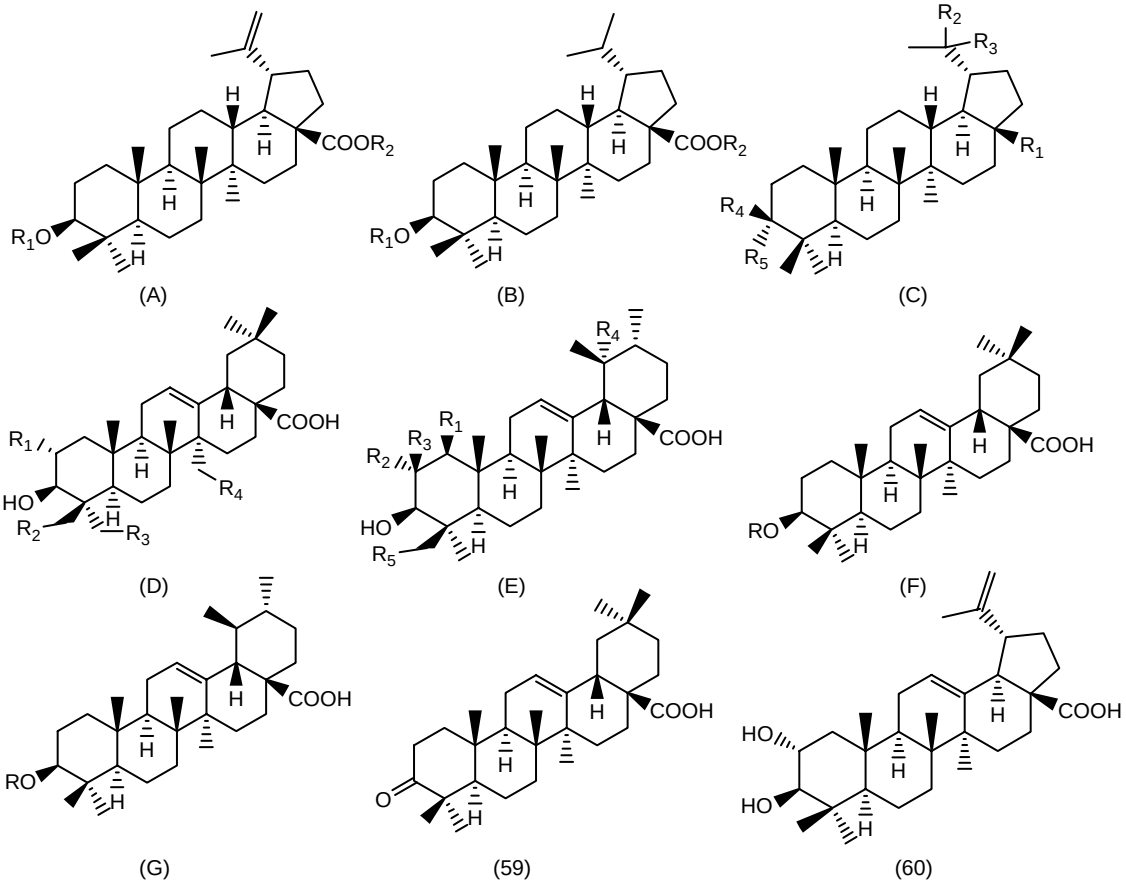
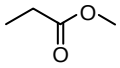
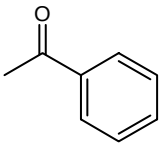
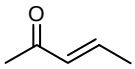
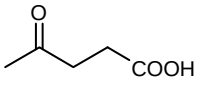
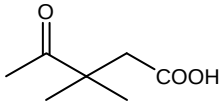
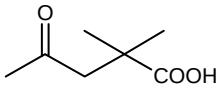
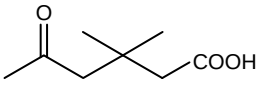
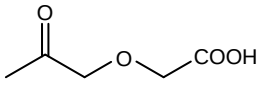
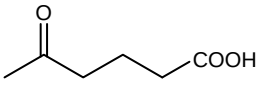
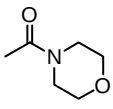


Figure 2. The triterpene skeletons of compounds (1)–(58) and the chemical structure of compounds (59)–(60).

Table 1. Substitutions for triterpene skeleton (A).

No.	R_1	R_2
1	H	
2		H
3		H
4		H
5		H
6	H	H
7		H
8		H
9		H
10		H
11		H
12	H	

subjected to a refined geometry optimisation using the AM1 semiempirical quantum MO-theory. The AM1 algorithm was selected because it gives good estimates of molecular energies and the computation time is much shorter than needed by *ab initio* methods. Finally, the optimised stereo conformation for compounds was obtained (the optimised stereo conformation of betulinic acid was shown on Figure 3 as an example), and in total 21 descriptors were calculated according with this stereo conformation: total energy (ET), binding energy (EB), isolated atomic energy (EI), electronic energy (EE), core–core interaction (EC), heat of formation (HF), dipole (DM), point-dipole (PD), molecular surface area (SM), molecular volume (VM), hydration energy (HE), *n*-octanol/water partition (LOGP), molar refractivity (RM), polarizability (PM), mass of molecule (MM), Highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO), as well as quantum chemical indices of electronegativity (χ): $(-0.5(\mathrm{HOMO} - \mathrm{LUMO}))$, hardness ($\eta$): $(0.5(\mathrm{HOMO} + \mathrm{LUMO}))$, softness (S): $(1/\eta)$, electrophilicity (ω): $(\chi^2/2\eta)$ were calculated according to the method proposed by Tanikaivelan et al. [21].

2.2 Methodology

The MLR was used to generate linear models between dependent variable and descriptors. Because of the many descriptors considered, a stepwise procedure combining the forward or backward algorithms was used to select the pertinent descriptors. The MLR was implemented in C programming language and run on a Pentium PC.

The ANN was used to generate a nonlinear model. A neural network is by definition a system of simple processing elements, called neurons, which are connected to a network by a set of weights. The knowledge of the neural network is encoded in the values of its weights; to determine the weight values, a set of examples is needed of the output relation to the inputs. The task of determining the weights from these examples is called training and is basically a conventional estimation problem. For this purpose, the feed-forward back-propagation network (BP network) was widely used. Back-propagation (BP) refers to the way the training is implemented and involves using a generalised delta rule. A 'learning' rate parameter influences the rate of weight and bias adjustment and is the basis of the BP algorithm. The set of input data is propagated through the network to give a prediction of the output. The error in the prediction is used to systematically update the weights based upon gradient information. The network is trained by altering the weights until the error between the training data outputs and the network predicted outputs are small enough. The structure of a typical three-layer BP network is shown in Figure 4. Circles in Figure 4 represent neurons. The typical architecture of a BP network consisted of one input layer, one or more hidden layers and one output layer, each neuron in one layer was inter-connected with all of its next layer's neurons, but neurons on the same layer were not connected with each other.

Table 2. Substitutions for triterpene skeleton (B).

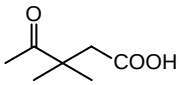
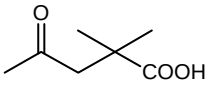
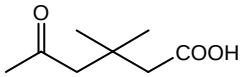
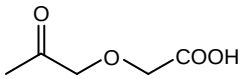
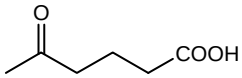
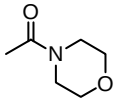
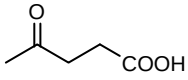
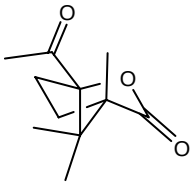
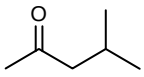
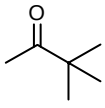
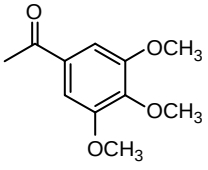
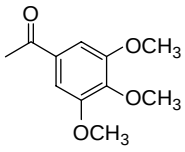
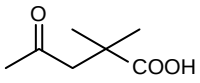
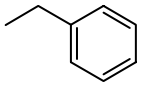
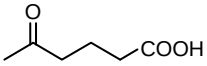
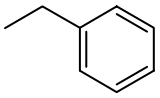
No.	R_1	R_2
13	H	H
14		H
15		H
16		H
17		H
18		H
19	H	
20		H
21		H
22		H
23		H
24		

Table 2 – *continued*

No.	R_1	R_2
25		
26		
27	H	

The neural networks used in this study were of the three-layer BP type with batch-gradient descent with a variable learning-rate algorithm [22]. The tan-sigmoid transfer function (tansig) was chosen for the neurons in the hidden layers and the linear transfer function (purelin) was used in the output layer. The neural network of BP algorithm was implemented in C programming language and run on a Pentium PC.

3. Results and analysis

3.1 Multiple linear regressions

MLR analysis was performed on the 60 triterpene derivatives. In order to obtain a reliable QSAR model, the available dataset should be divided into a training and

Table 3. Substitutions for triterpene skeleton (C).

No	R_1	R_2	R_3	R_4	R_5
28	COOH	O		OH	H
29	CH_2OH	CH_2		OH	H
30	COOH	CH_3	H	O	
31	$COOCON(CH_3)_2$	CH_3	H	O	
32	COOH	CH_2		O	

Table 4. Substitutions for triterpene skeleton (D).

No	R_1	R_2	R_3	R_4
33	H	H	H	H
34	OH	H	OH	H
35	OH	OH	H	H
36	H	H	H	OH

Table 5. Substitutions for triterpene skeleton (E).

No.	R_1	R_2	R_3	R_4	R_5
37	H	H	H	OH	H
38	H	H	H	H	H
39	H	OH	H	H	H
40	H	OH	H	H	OH
41	H	OH	H	OH	H
42	H	O		OH	H
43	OH	O		OH	H

a prediction set. The training set was used to build linear models so that an accurate relationship could be found between structure and biological activity. The prediction set is a group of molecules not used to develop the model and that serve to test the predictive ability of the model with unknown compounds. The date set was divided into training set and prediction set (10%) by clustering algorithm proposed by Burden [23].

With the PIC_{50} as the dependent variable and all descriptors as the independent variables, a stepwise procedure was used to build a linear model. The equation was as following:

$$PIC_{50} = -7.0 \times 10^{-5}\,EB - 0.001\,HE - 0.23\,LOGP - 1.525\,HOMO + 0.093\omega. \quad (1)$$

This linear model showed a very low correlation coefficient (r) of 0.4567 for training set and 0.3945 for prediction set. The results showed that the relationship between computational descriptors and experimental anti-HIV activity of triterpene derivatives could not be explained with a simple linear model.

Table 6. Substitutions for triterpene skeleton (F).

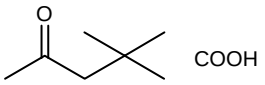
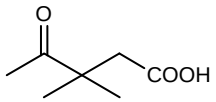
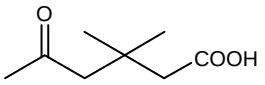
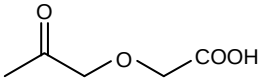
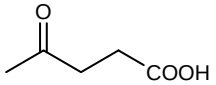
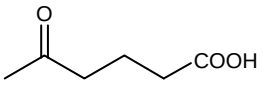
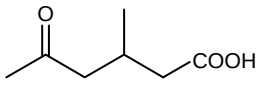
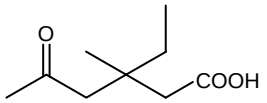
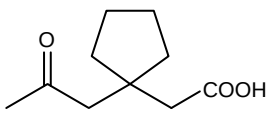
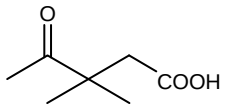
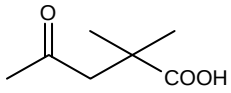
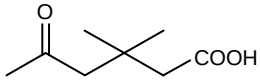
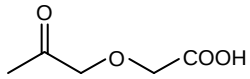
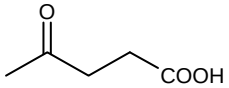
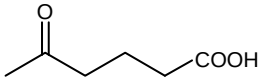
No.	R
44	
45	
46	
47	
48	
49	
50	
51	
52	

Table 7. Substitutions for triterpene skeleton (G).

No.	R
53	
54	
55	
56	
57	
58	

3.2 *Artificial neural network*

The ANN approach is especially suitable for analysing complex nonlinear relationships between the output and inputs. The major advantage of ANN lies in the fact that

Table 8. The activity value of 60 triterpene derivatives.

No.	IC_{50}	PIC_{50}	Calculated PIC_{50}	No.	IC_{50}	PIC_{50}	Calculated PIC_{50}
1	26	1.41	1.39	31[a]	13.3	1.12	1.18
2[b]	20	1.30	1.43	32	1.8	0.26	0.14
3	15	1.18	1.21	33	47.8	1.68	1.49
4	48	1.68	1.34	34	16.4	1.21	1.58
5	16	1.20	1.27	35	73.7	1.87	1.65
6	13.0	1.11	1.10	36	52.9	1.72	1.93
7[b]	15.9	1.20	0.99	37[a]	48.9	1.69	1.50
8	7	0.85	0.89	38[b]	14.3	1.16	1.13
9	4.5	0.65	0.74	39	16.9	1.23	1.25
10	11.7	1.07	1.22	40	18.4	1.26	1.50
11	12.8	1.11	1.19	41	30.7	1.49	1.30
12	11.4	1.06	0.99	42[a]	82.2	1.91	2.05
13	12.6	1.10	1.07	43	79.6	1.90	2.16
14[a]	7.7	0.89	0.71	44	19.2	1.28	1.16
15	4.9	0.69	0.89	45	16.6	1.22	1.21
16	5.8	0.76	0.54	46	23.7	1.37	1.10
17	13.1	1.12	1.22	47	38.6	1.59	1.48
18	7.9	0.90	1.03	48	11.7	1.07	1.06
19[a]	35	1.54	1.45	49[b]	25.6	1.41	1.53
20	13.4	1.13	1.30	50	47.0	1.67	1.40
21	1	0.00	0.12	51	7.2	0.86	0.83
22	83	1.92	1.87	52	7.5	0.88	0.84
23	83	1.92	1.52	53	30.7	1.49	1.32
24	53	1.72	1.76	54	49.5	1.70	1.33
25[b]	13	1.11	1.04	55	7.00	0.85	1.29
26[a]	9.3	0.97	0.80	56	48.2	1.68	1.52
27	6.6	0.82	1.34	57	44.9	1.65	1.83
28	90	1.95	1.84	58[b]	30.7	1.49	1.61
29	45	1.65	1.58	59	1.28	0.11	0.27
30	0.032	−1.49	−1.12	60	8.50	0.93	0.97

[a] Prediction set for ANN model. [b] Validation set for ANN model.

QSAR can be developed without having to specify the analytical form of a particular correlation model.

Descriptors should be selected to serve as the inputs data for ANN model because inputs data of ANN model with highly correlated or contained redundant information would greatly increase the complexity of the model and would be bad for the accuracy of the model, which was carried out by following steps:

(1) Correlation coefficients (r) were calculated for all descriptors. The results are listed in Table 9.
(2) According to the correlation matrix, descriptors were partitioned into several groups with descriptors in one group having a correlation coefficient value greater than a user specified cutoff (0.85 in this case). In this study, ten groups were obtained: (1) ET, EB, EI, EE, EC, SM, VM, PM, MM, RM; (2) HF; (3) DM, PD; (4) HE; (5) LOGP; (6) HOMO; (7) LUMO; (8) χ, ω; (9) η; (10) S.
(3) Sensitivity analysis of descriptors was used to select descriptors from each group. The correlation coefficients between descriptors and dependent variable were determined. The higher the correlation coefficient of descriptor, the greater influences on the model: hence, those descriptors with higher correlation coefficients for each group were selected. Finally, five descriptors EB, LOGP, HOMO HE and ω were selected to serve as the inputs data for ANN model.

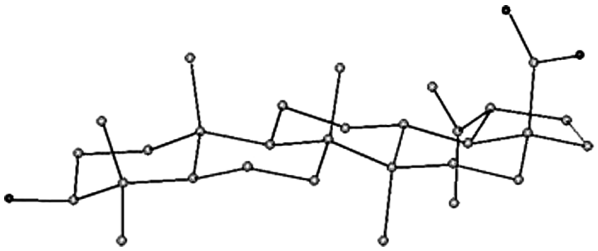


Figure 3. The optimised stereo conformation of betulinic acid.

In order to obtain a reliable QSAR model, the 60 molecules were divided into three sets: a training, a validation and a prediction set. The training set was used to build the model and the prediction set was used to test the predictive ability of the model with unknown compounds. The validation set was used to during the development of ANN models to prevent neural network over fitting. These sets were created using the clustering algorithm as before [23]. Ultimately, the training set contained 48 molecules, the validation set contained six molecules, and the prediction set contained six molecules (Table 8).

The number of neurons in the hidden layer is another important factor determining the network's performance. Different amounts of hidden neurons were studied, and the performance of the ANN model was evaluated with the combination of the root mean square (RMS) error for the training set, validation set and prediction set. The results are summarised in Table 10. The model architecture of 5-6-1 with minimum RMS of training,

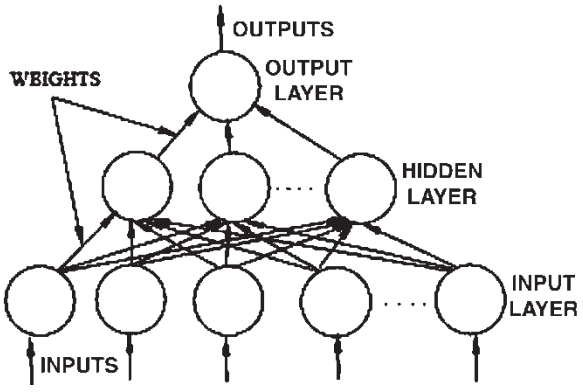


Figure 4. Structure of BP network.

Table 9. Correlation matrix (r value) of descriptors.

	ET	EB	EI	EE	EC	HF	DM	PD	SM	VM	HE	LOG-P	RM	PM	MM	HOMO	LUMO	χ	η	S
EB	0.97	1.00																		
EI	1.00	0.96	1.00																	
EE	0.98	0.98	0.98	1.00																
EC	−0.98	−0.98	−0.98	−1.00	1.00															
HF	0.82	0.72	0.82	0.73	−0.72	1.00														
DM	−0.12	−0.03	−0.13	−0.10	0.10	−0.09	1.00													
PD	0.01	0.07	0.01	0.01	−0.01	0.09	0.91	1.00												
SM	−0.95	−0.95	−0.95	−0.94	0.94	−0.73	0.02	−0.07	1.00											
VM	−0.95	−0.97	−0.95	−0.96	0.96	−0.71	0.02	−0.07	0.99	1.00										
HE	−0.74	−0.73	−0.74	−0.68	0.68	−0.72	0.13	0.00	0.80	0.78	1.00									
LOGP	−0.57	−0.73	−0.56	−0.64	0.64	−0.27	−0.18	−0.21	0.72	0.75	0.65	1.00								
RM	−0.95	−0.98	−0.95	−0.97	0.97	−0.67	0.04	−0.05	0.98	0.99	0.73	0.75	1.00							
PM	−0.95	−0.98	−0.95	−0.97	0.97	−0.67	0.04	−0.06	0.98	0.99	0.74	0.76	1.00	1.00						
MM	−0.98	−0.97	−0.98	−0.98	0.98	−0.75	0.09	−0.02	0.98	0.99	0.76	0.68	0.99	0.99	1.00					
HOMO	0.10	0.19	0.09	0.13	−0.13	0.16	0.00	−0.09	−0.16	−0.17	−0.29	−0.37	−0.12	−0.14	−0.12	1.00				
LUMO	0.48	0.44	0.48	0.50	−0.50	0.07	−0.25	−0.18	−0.46	−0.46	−0.23	−0.29	−0.50	−0.49	−0.48	−0.35	1.00			
χ	0.19	0.11	0.20	0.18	−0.18	−0.07	−0.14	−0.04	−0.14	−0.14	0.07	0.10	−0.19	−0.17	−0.18	−0.86	0.77	1.00		
η	0.46	0.52	0.46	0.51	−0.51	0.21	−0.19	−0.22	−0.51	−0.52	−0.46	−0.58	−0.50	−0.51	−0.49	0.69	0.43	−0.24	1.00	
S	−0.46	−0.52	−0.46	−0.50	0.51	−0.23	0.19	0.22	0.51	0.52	0.48	0.59	0.49	0.50	0.48	−0.73	−0.39	0.28	−1.00	1.00
ω	−0.04	0.06	−0.04	−0.01	0.01	0.09	0.05	−0.05	−0.03	−0.04	−0.21	−0.29	0.02	0.00	0.02	0.97	−0.48	−0.92	0.57	−0.61

Table 10. The influence of different hidden neuron on the ANNs performance.

ANNs*	Training set RMS	Validation set RMS	Prediction set RMS
5-4-1	0.3407	0.3407	0.4835
5-5-1	0.3209	0.3658	0.4852
5-6-1	0.2019	0.2214	0.2883
5-7-1	0.2966	0.4098	0.3451
5-8-1	0.3312	0.3176	0.4694

*Number of inputs-hidden-outputs neurons.

validation and prediction set was 0.2019, 0.2214 and 0.2883, respectively. The architecture of 5-6-1 was selected as the final model architecture.

In order to make the training more efficient, inputs and outputs were normalised so that they fell in the range of [−1,1] before training. After the network has been trained, the outputs are to be transferred back to the same units that were used for the original outputs for comparison purpose.

A final BP ANN model of 5-6-1 architecture, that is one input layer with five neurons (EB, LOGP, HOMO, HE and ω), one hidden layer with six neurons and one output layer with one neuron (PIC_{50}) was at last optimal. The parameters of the final BP ANN model were as following: initial weights were random values between 0 and 1, the initial learning rate was 0.020, the increased learning rate was 1.05, the goal of stopping cycle MSE was 0.05, the maximum overlapping cycles were 4000. The calculated PIC_{50} from the final BP ANN model is listed in Table 8 and plots of calculated PIC_{50} versus experimental PIC_{50} are shown in Figure 5. The correlation coefficients (r) and RMS of training, validation and prediction set for the final BP ANN

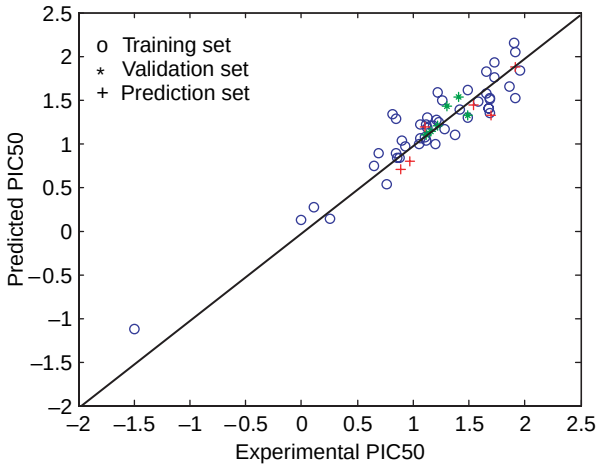


Figure 5. A plot of experimental versus predicted PIC50 produced from the nonlinear model using a 5-6-1 architecture.

Table 11. The r and RMS for ANN model*.

Training set		Validation set		Prediction set	
r	RMS	r	RMS	r	RMS
0.8895	0.2019	0.8843	0.2214	0.8819	0.2883

*The ANN architecture of 5-6-1.

model are listed in Table 11. It can be seen clearly from Figure 5 and Table 11 that the results of ANN were satisfied.

In order to ensure that the results obtained in ANN were not due to chance and to lend credence to our results, we ran a scrambling experiment. The dependent variable PIC_{50} was scrambled randomly, then the same algorithms used in ANN again. The statistical results as the RMS error of its results were compared with that of the final BP ANN model. The ranging of RMS was from 0.012 to 0.345. This indicated chance correlation does not play a significant role in the results described above.

In this study, we failed to generate a QSAR model with MLR for the correlation coefficient (r) below 0.5 for both training set and prediction set. The QSAR model with ANN performed well, the r of training, validation and prediction set are all above 0.88, and the RMS for the three sets are all below 0.3. As can be seen from Figure 4, the points were spread along with the line closely, which indicated that the ANN model performed well for both training and prediction. The reason the nonlinear model performed well was likely due to the complex relationship between the structure of compounds and the action mechanism.

4. Conclusions

The MLR and ANN were employed to study the model and prediction of PIC_{50} value from computationally derived molecular descriptors of triterpene derivatives. The pattern obtained with the ANN approach showed ANN was more efficient than MLR analysis, reflecting the fact that the relationship between descriptors and biological activity of triterpene derivatives is nonlinear. The MLRs showed a poor result which was due to the risk of including noise in the model while trying to account for the nonlinearity.

In this study, total 21 descriptors were employed to develop a QSAR model. Different methods were used to select the most relevant descriptors for MLR and ANN. The stepwise procedure combining the forward or backward algorithms was used in MLR and correlation matrix combining sensitivity analysis was used in ANN to select the most relevant descriptor. Finally, five descriptors, EB, LOGP, HOMO, HE and ω, were selected for both MLR and ANN. These descriptors are

thus playing an important role in the anti-HIV activity of triterpene derivatives with diversity structures.

The results of this study revealed good stability of the structure–activity relationship, which offers guidance to the further design of new series of triterpene derivatives.

Acknowledgements

This work was supported in part by the West China Center of Medical Sciences.

References

[1] D. Boden, A. Hurley, L. Zhang, Y. Cao, Y. Guo, E. Jones, J. Tsay, J. Ip, C. Farthing, K. Limoli, N. Parkin, and M. Markowitz, *HIV-1 drug resistance in newly infected individuals*, J. Am. Med. Assoc. 282 (1999), pp. 1135–1141.

[2] M. Louie and M. Markowitz, *Goals and milestones during treatment of HIV-1 infection with antiretroviral therapy: A pathogenesis-based perspective*, Antiviral Res. 55 (2002), pp. 15–25.

[3] Y. Kashiwada, F. Hashimoto, L.M. Cosentino, C.-H. Chen, P.E. Garrett, and K.-H. Lee, *Betulinic acid and dihydrobetulinic acid derivatives as potent anti-HIV agents*, J. Med. Chem. 39 (1996), pp. 1016–1017.

[4] C. Aiken and C.H. Chen, *Betulinic acid derivatives as HIV-1 antivirals*, Trends Mol. Med. 11 (2005), pp. 31–36.

[5] F. Soler, C. le Poujade, M. Evers, J.-C. Carry, Y. Henin, A. Bousseau, T. Huet, R. Pauwels, E. De Clercq, J.-F. Mayaux et al., *Betulinic acid derivatives: A new class of specific inhibitors of human immunodeficiency virus type 1 entry*, J. Med. Chem. 39 (1996), pp. 1069–1083.

[6] I.-C. Sun, C.-H. Chen, Y. Kashiwada, J.-H. Wu, H.-K. Wang, and K.-H. Lee, *Anti-AIDS agents 49.1 synthesis, anti-HIV, and anti-fusion activities of IC9564 analogues based on betulinic acid*, J. Med. Chem. 45 (2002), pp. 4271–4275.

[7] M. Evers, C. le Poujade, F. Soler, Y. Ribeill, C. James, Y. Lelievre, J.-C. Gueguen, D. Reisdorf, I. Morize, R. Pauwels et al., *Betulinic acid derivatives: A new class of human immunodeficiency virus type 1 specific inhibitors with a new mode of action*, J. Med. Chem. 39 (1996), pp. 1056–1068.

[8] L. Huang, X. Yuan, C. Aiken, and C.-H. Chen, *Bifunctional anti-human immunodeficiency virus type 1 small molecules with two novel mechanisms of action*, Antimicrob. Agents Chemother. 48 (2004), pp. 663–665.

[9] L. Huang, P. Ho, K.-H. Lee, and C.-H. Chen, *Synthesis and anti-HIV activity of bi-functional betulinic acid derivatives*, Bioorg. Med. Chem. 14 (2006), pp. 2279–2289.

[10] B. Labrosse, C. Treboute, and M. Alizon, *Sensitivity to a nonpeptidic compound (RPR103611) blocking human immunodeficiency virus type 1 env-mediated fusion depends on sequence and accessibility of the gp41 loop region*, J. Virol. 74 (2000), pp. 2142–2150.

[11] S.L. Holz-Smith, I.C. Sun, L. Jin, T.J. Mathews, K.H. Lee, and C.H. Chen, *Role of human immunodeficiency virus (HIV) type 1 envelop in the anti-HIV activity of the betulinic acid derivative IC9564*, Antimicrob. Agents Chemother. 45 (2001), pp. 60–66.

[12] T. Pengsuparp, L. Cai, H. Constant, H.H.S. Fong, L.Z. Lin, A.D. Kinghorn, J.M. Pezzuto, G.A. Cordell, K. Ingolfsdottir, H. Wagner et al., *Mechanistic evaluation of new plant-derived compounds that inhibit HIV-1 reverse transcriptase*, J. Nat. Prod. 58 (1995), pp. 1024–1031.

[13] T. Pengsuparp, L. Cai, H.H.S. Fong, A.D. Kinghorn, J.M. Pezzuto, M.C. Wani, and M.E. Wall, *Pentacyclic triterpenes derived from* Maprounea africana *are potent inhibitors of HIV-1 reverse transcriptase*, J. Nat. Prod. 57 (1994), pp. 415–418.

[14] L. Quere, T. Wenger, and H.J. Schramm, *Triterpenes as potential dimerization inhibitors of HIV-1 protease*, Biochem. Biophys. Res. Commun. 227 (1996), pp. 484–488.

[15] H.X. Xu, D.S. Ming, H. Dong, and P.P. But, *A new anti-HIV triterpene from* Geum japonicum, Chem. Pharm. Bull. 48 (2000), pp. 1367–1369.

[16] D. Yu, Y. Sakurai, C.H. Chen, F.-R. Chang, L. Huang, Y. Kashiwada, K.H. Lee, *Anti-AIDS agents 69. Moronic acid and other triterpene derivatives as novel potent anti-HIV agents*, J. Med. Chem. 49 (2006), pp. 5462–5469.

[17] T. Fujioka, Y. Kashiwada, R.E. Kilkuskle, L.M. Cosentino, L.M. Ballas, J.B. Jiang, W.P. Janzen, I.S. Chen, and K.H. Lee, *Anti-AIDS agents, 11. Betulinic acid and platanic acid as anti-HIV principles from* Syzigium claviflorum*, and the anti-HIV activity of structurally related triterpenoids*, J. Nat. Prod. 57 (1994), pp. 243–247.

[18] F. Hashimoto, Y. Kashiwada, L.M. Cosentino, C.-H. Chen, P.E. Garrett, and K.-H. Lee, *Anti-AIDS agents-XXVII synthesis and anti-HIV activity of betulinic acid and dihydrobetulinic acid derivatives*, Bioorg. Med. Chem. 5 (1997), pp. 2133–2143.

[19] Y. Kashiwada, H.-K. Wang, T. Nagao, S. Kitanaka, I. Yasuda, T. Fujioka, T. Yamagishi, L.M. Cosentino, M. Kozuka, H. Okabe et al., *Anti-AIDS agents 30. Anti-HIV activity of oleanolic acid, pomolic acid, and structurally related triterpenoids*, J. Nat. Prod. 61 (1998), pp. 1090–1095.

[20] Y. Kashiwada, T. Nagao, A. Hashimoto, Y. Ikeshiro, H. Okabe, L.M. Cosentino, and K.-H. Lee, *Anti-AIDS agents 38. Anti-HIV activity of 3-O-acyl Ursolic acid derivatives*, J. Nat. Prod. 63 (2000), pp. 1619–1622.

[21] P. Thanikaivelan, V. Subramanian, J.R. Rao, and B.U. Nair, *Application of quantum chemical descriptor in quantitative structure activity and structure property relationship*, Chem. Phys. Lett. 323 (2000), pp. 59–70.

[22] M.T. Hagan, H.B. Demuth, and M.H. Beale, *Neural Network Design*, Boston (MA), PWS Publishing, 1996.

[23] F.R. Burden and D.A. Winkler, *Robust QSAR models using Bayesian regularized neural networks*, J. Med. Chem. 42 (1999), pp. 3183–3187.